MULTIVARIATE DENSITY ESTIMATION UNDER SUP-NORM LOSS: ORACLE APPROACH, ADAPTATION AND INDEPENDENCE STRUCTURE

Oleg Lepski

The Annals of Statistics
2013, Vol. 41, No. 2, 1005–1034
DOI: 10.1214/13-AOS1109

MULTIVARIATE DENSITY ESTIMATION UNDER SUP-NORM LOSS: ORACLE APPROACH, ADAPTATION AND INDEPENDENCE STRUCTURE

BY OLEG LEPSKI

Université Aix-Marseille

This paper deals with the density estimation on $\mathbb{R}^d$ under sup-norm loss. We provide a fully data-driven estimation procedure and establish for it a so-called sup-norm oracle inequality. The proposed estimator allows us to take into account not only approximation properties of the underlying density, but eventual independence structure as well. Our results contain, as a particular case, the complete solution of the bandwidth selection problem in the multivariate density model. Usefulness of the developed approach is illustrated by application to adaptive estimation over anisotropic Nikolskii classes.

1. Introduction. Let $(\Omega, \mathfrak{A}, \mathrm{P})$ be a complete probability space, and let $X_i = (X_{1,i}, \dots, X_{d,i})$, $i \geq 1$, be the sequence of $\mathbb{R}^d$-valued i.i.d. random variables defined on $(\Omega, \mathfrak{A}, \mathrm{P})$ and having the density f with respect to lebesgue measure. Furthermore, $\mathbb{P}_f^{(n)}$ denotes the probability law of $X^{(n)} = (X_1, \dots, X_n)$, $n \in \mathbb{N}^*$ and $\mathbb{E}_f^{(n)}$ is the mathematical expectation with respect to $\mathbb{P}_f^{(n)}$.

The objective is to estimate the density f and the quality of any estimation procedure, that is, $X^{(n)}$-measurable mapping $\widehat{f}_n : \mathbb{R}^d \to \mathbb{L}_\infty(\mathbb{R}^d)$, is measured by *sup-norm risk* given by

$$R_n^{(q)}(\widehat{f}, f) = \left(\mathbb{E}_f^{(n)} \|\widehat{f}_n - f\|_\infty^q\right)^{1/q}, \qquad q \geq 1.$$

It is well known that even asymptotically ($n \to \infty$) the quality of estimation given by $R_n^{(q)}$ heavily depends on the dimension d. However, these asymptotics can be essentially improved if the underlying density possesses some special structure. Let us briefly discuss one of these possibilities which will be exploited in the sequel.

Introduce the following notation. Let $\mathcal{I}_d$ be the set of all subsets of $\{1, \dots, d\}$. For any $\mathbf{I} \in \mathcal{I}_d$ denote $x_\mathbf{I} = \{x_j \in \mathbb{R}, j \in \mathbf{I}\}$, $\bar{\mathbf{I}} = \{1, \dots, d\} \setminus \mathbf{I}$ and let $|\mathbf{I}| = \mathrm{card}(\mathbf{I})$. Moreover for any function $g : \mathbb{R}^{|\mathbf{I}|} \to \mathbb{R}$ we denote $\|g\|_{\mathbf{I},\infty} = \sup_{x_\mathbf{I} \in \mathbb{R}^{|\mathbf{I}|}} |g(x_\mathbf{I})|$. Define also

$$f_\mathbf{I}(x_\mathbf{I}) = \int_{\mathbb{R}^{|\bar{\mathbf{I}}|}} f(x)\,\mathrm{d}x_{\bar{\mathbf{I}}}, \qquad x_\mathbf{I} \in \mathbb{R}^{|\mathbf{I}|}.$$

Received June 2012; revised January 2013.
MSC2010 subject classifications. 62G07.
Key words and phrases. Density estimation, oracle inequality, adaptation, upper function.

In accordance with this definition we put $f_{\mathbf{I}} \equiv 1, \mathbf{I} = \varnothing$. Note that $f_{\mathbf{I}}$ is the marginal density of $X_{\mathbf{I},1} := \{X_{j,1}, j \in \mathbf{I}\}$. Denote by $\mathfrak{P}$ the set of all partitions of $\{1, \dots, d\}$ completed by empty set $\varnothing$, and we will use $\bar{\varnothing}$ for $\{1, \dots, d\}$. For any density f let

$$\mathfrak{P}(f) = \left\{ \mathcal{P} \in \mathfrak{P} : f(x) = \prod_{\mathbf{I} \in \mathcal{P}} f_{\mathbf{I}}(x_{\mathbf{I}}), \forall x \in \mathbb{R}^d \right\}.$$

First we note that $f \equiv f_{\bar{\varnothing}}$, and therefore $\mathfrak{P}(f)$ is not empty since $\bar{\varnothing} \in \mathfrak{P}(f)$ for any f. Next, if $\mathcal{P} \in \mathfrak{P}(f)$, then $\{X_{\mathbf{I},1}\mathbf{I} \in \mathcal{P}\}$ are independent random vectors. At last, if $X_{1,1}, \dots, X_{d,1}$, are independent random variables, then obviously $\mathfrak{P}(f) = \mathfrak{P}$.

Suppose now that there exists $\boldsymbol{\mathcal{P}} \neq \bar{\varnothing}$ such that $\boldsymbol{\mathcal{P}} \in \mathfrak{P}(f)$. If this partition is known, we can proceed as follows. For any $\mathbf{I} \in \boldsymbol{\mathcal{P}}$ basing on observation $X_{\mathbf{I}}^{(n)}$, we estimate first the marginal density $f_{\mathbf{I}}$ by $\widehat{f}_{\mathbf{I},n}$ and then construct the estimator for joint density f as

$$\widehat{f}_n(x) = \prod_{\mathbf{I} \in \boldsymbol{\mathcal{P}}} \widehat{f}_{\mathbf{I},n}(x_{\mathbf{I}}).$$

One can expect (and we will see that our conjecture is true) that quality of estimation provided by this estimator will correspond not to the dimension d but to so-called effective dimension, which in our case is defined as $d(\boldsymbol{\mathcal{P}}) = \sup_{\mathbf{I} \in \boldsymbol{\mathcal{P}}} |\mathbf{I}|$. The main difficulty we meet trying to realize the latter construction is that the knowledge of $\boldsymbol{\mathcal{P}}$ is not available. Moreover, our structural hypothesis cannot be true in general, that is expressed formally by $\mathfrak{P}(f) = \{\bar{\varnothing}\}$. So, one of the problem we address in the present paper consists in adaptation to unknown configuration $\boldsymbol{\mathcal{P}} \in \mathfrak{P}(f)$.

We note, however, that even if $\boldsymbol{\mathcal{P}}$ is known, for instance, $\boldsymbol{\mathcal{P}} = \bar{\varnothing}$, the quality of an estimation procedure depends often on approximation properties of f or $\{\widehat{f}_{\mathbf{I},n}, \mathbf{I} \in \boldsymbol{\mathcal{P}}\}$. So, our second goal is to construct an estimator which would mimic an estimator corresponding to the minimal, and therefore unknown, approximation error. Using modern statistical language our goal here is to mimic an oracle. It is important to emphasize that we would like to solve both aforementioned problems simultaneously. Let us now proceed with detailed consideration.

Collection of estimators. Let $\mathbf{K} : \mathbb{R} \to \mathbb{R}$ be a given function satisfying the following assumption.

ASSUMPTION 1. $\int \mathbf{K} = 1$, $\|\mathbf{K}\|_\infty < \infty$, $\mathrm{supp}(\mathbf{K}) \subseteq [-1/2, 1/2]$, $\mathbf{K}$ is symmetric and

$$\exists L > 0: \qquad \left|\mathbf{K}(t) - \mathbf{K}(s)\right| \le L|t - s| \qquad \forall t, s \in \mathbb{R}.$$

Put for $\mathbf{I} \in \mathcal{I}_d$

$$K_{h_{\mathbf{I}}}(u) = V_{h_{\mathbf{I}}}^{-1} \prod_{j \in \mathbf{I}} \mathbf{K}(u_j / h_j), \qquad V_{h_{\mathbf{I}}} = \prod_{j \in \mathbf{I}} h_j.$$

For two vectors u, v here and later u/v denotes coordinate-vise division. We will use the notation $V_h = \prod_{j=1}^{d} h_j$ instead of $V_{h_{\mathbf{I}}}$ when $\mathbf{I} = \{1, \dots, d\}$. Denote also $\mathrm{k}_m = \|\mathbf{K}\|_m$, $m = \{1, \infty\}$.

For any $p \ge 1$ let $\gamma_p : \mathbb{N}^* \times \mathbb{R}_+ \to \mathbb{R}_+$ be the function whose explicit expression is given in Section 2.3 (its expression is quite cumbersome, and it is not convenient for us to present it right now).

Introduce the notation (remember that q is the quantity involved in the definition of the risk)

$$\mathcal{H}_n = \{h \in (0, 1]^d : n V_h \ge (\mathfrak{a}^*)^{-1} \ln(n)\}, \qquad \mathfrak{a}^* = \inf_{\mathbf{I} \in \mathcal{I}_d} [2\gamma_{2q}(|\mathbf{I}|, \mathrm{k}_\infty)]^{-2},$$

and for any $\mathbf{I} \in \mathcal{I}_d$ and $h \in \mathcal{H}_n$ consider the kernel estimator

$$\widetilde{f}_{h_{\mathbf{I}}}(x_{\mathbf{I}}) = n^{-1} \sum_{i=1}^{n} K_{h_{\mathbf{I}}}(X_{\mathbf{I},i} - x_{\mathbf{I}}).$$

Introduce the family of estimators

$$\mathfrak{F}(\mathfrak{P}) = \left\{ \widehat{f}_{h,\mathcal{P}}(x) = \prod_{\mathbf{I} \in \mathcal{P}} \widetilde{f}_{h_{\mathbf{I}}}(x_{\mathbf{I}}), x \in \mathbb{R}^d, \mathcal{P} \in \mathfrak{P}, h \in \mathcal{H}_n \right\}.$$

In particular, $\widehat{f}_{h,\bar{\varnothing}}(x) = n^{-1} \sum_{i=1}^{n} K_h(X_i - x)$, $x \in \mathbb{R}^d$, is the Parzen–Rosenblatt estimator [Parzen (1962), Rosenblatt (1956)] with kernel K and multi-bandwidth h. Our goal is to propose a data-driven selection from the family $\mathfrak{F}(\mathfrak{P})$.

The estimation of a probability density is the subject of the vast literature. We do not pretend here to provide a complete overview and only present the results relevant in the context of the considered problems. Minimax and minimax adaptive density estimation with $\mathbb{L}_s$-risks were considered in Bretagnolle and Huber (1979), Ibragimov and Khasminskii (1980, 1981), Devroye and Györfi (1985), Efroimovich (1986, 2008), Khasminskii and Ibragimov (1990), Donoho et al. (1996), Golubev (1992), Kerkyacharian, Picard and Tribouley (1996), Juditsky and Lambert-Lacroix (2004), Rigollet (2006), Mason (2009), Reynaud-Bouret, Rivoirard and Tuleau-Malot (2011) and Akakpo (2012), where further references can be found. Oracle inequalities for $\mathbb{L}_s$-risks for $s = 1$ and $s = 2$ were established in Devroye and Lugosi (1996, 1997, 2001), Massart (2007), Chapter 7, Samarov and Tsybakov (2007), Rigollet and Tsybakov (2007) and Birgé (2008). The last cited paper contains a detailed discussion of recent developments in this area. The bandwidth selection problem in the density estimation on $\mathbb{R}^d$ with $\mathbb{L}_s$-risks for any $1 \le s < \infty$ was studied in Goldenshluger and Lepski (2011). The oracle inequalities obtained there were used for deriving adaptive minimax results over the collection of anisotropic Nikolskii classes.

The adaptive estimation under sup-norm loss was initiated in Lepskiĭ (1991, 1992) and continued in Tsybakov (1998) in the framework of Gaussian white noise model. Then it was developed for anisotropic functional classes in Bertin (2005).

The adaptive estimation of a probability density on $\mathbb{R}$ in sup-norm was the subject of recent papers Giné and Nickl (2009, 2010) and Gach, Nickl and Spokoiny (2013).

Organization of the paper. In Section 2 we present a data-driven selection procedure from $\mathfrak{F}(\mathfrak{P})$ and establish for it sup-norm oracle inequality. Section 3 is devoted to the adaptive estimation over the collection of anisotropic Nikolskii classes of functions. The proof of main results are given in Section 4, and technical lemmas are proven in the Appendix.

2. Oracle inequality. Let $\mathcal{P} \in \mathfrak{P}$ be fixed and define for any $h, \eta \in \mathcal{H}_n$ and any $\mathbf{I} \in \mathcal{P}$

$$\widetilde{f}_{h_{\mathbf{I}},\eta_{\mathbf{I}}}(x_{\mathbf{I}}) = n^{-1} \sum_{i=1}^{n} [K_{h_{\mathbf{I}}} \star K_{\eta_{\mathbf{I}}}](X_{\mathbf{I},i} - x_{\mathbf{I}}), \tag{2.1}$$

where $[K_{h_{\mathbf{I}}} \star K_{\eta_{\mathbf{I}}}] = \prod_{j \in \mathbf{I}} [\mathbf{K}_{h_j} * \mathbf{K}_{\eta_j}]$ and $[\mathbf{K}_{h_j} * \mathbf{K}_{\eta_j}](z) = \int_{\mathbb{R}} \mathbf{K}_{h_j}(u - z) \times \mathbf{K}_{\eta_j}(u)\, \mathrm{d}u$, $z \in \mathbb{R}$.

Note that "$\star$" is the convolution operator on $\mathbb{R}^{|\mathbf{I}|}$. Define

$$\mathbf{f}_n = \sup_{h \in \mathcal{H}_n} \sup_{\mathbf{I} \in \mathcal{I}_d} \left\| n^{-1} \sum_{i=1}^{n} \left| K_{h_{\mathbf{I}}}(X_{\mathbf{I},i} - \cdot) \right| \right\|_{\mathbf{I},\infty}, \qquad \bar{\mathbf{f}}_n = 1 \vee 2\mathbf{f}_n,$$

$$\widehat{A}_n(h, \mathcal{P}) = \sqrt{\frac{\bar{\mathbf{f}}_n \ln(n)}{n V(h, \mathcal{P})}}, \qquad V(h, \mathcal{P}) = \inf_{\mathbf{I} \in \mathcal{P}} V_{h_{\mathbf{I}}}.$$

Let us endow the set $\mathfrak{P}$ with the operation "$\diamond$" putting for any $\mathcal{P}, \mathcal{P}' \in \mathfrak{P}$

$$\mathcal{P} \diamond \mathcal{P}' = \{\mathbf{I} \cap \mathbf{I}' \neq \varnothing, \mathbf{I} \in \mathcal{P}, \mathbf{I}' \in \mathcal{P}'\} \in \mathfrak{P}.$$

Introduce for any $h, \eta \in \mathcal{H}_n$ and any $\mathcal{P}, \mathcal{P}'$ the estimator

$$\widehat{f}_{(h,\mathcal{P}),(\eta,\mathcal{P}')}(x) = \prod_{\mathbf{I}^{\diamond} \in \mathcal{P} \diamond \mathcal{P}'} \widetilde{f}_{h_{\mathbf{I}^{\diamond}},\eta_{\mathbf{I}^{\diamond}}}(x_{\mathbf{I}^{\diamond}}), \qquad x \in \mathbb{R}^d.$$

Set finally $\Lambda = \sup_{\mathcal{P} \in \mathfrak{P}} \sup_{\mathbf{I} \in \mathcal{P}} \gamma_{2q}(|\mathbf{I}|, \mathrm{k}_\infty)$, and let $\lambda = \Lambda d (\bar{\mathbf{f}}_n)^{d^2/4}$.

2.1. *Selection procedure*. Let $\overline{\mathfrak{P}} \subseteq \mathfrak{P}$, satisfying $\bar{\varnothing} = \{1, \dots, d\} \in \overline{\mathfrak{P}}$, and $\overline{\mathcal{H}}_n \subseteq \mathcal{H}_n$ be fixed. Without further mentioning, we will assume that either $\overline{\mathcal{H}}_n = \mathcal{H}_n$ or $\overline{\mathcal{H}}_n$ is finite.

For any $\mathcal{P} \in \overline{\mathfrak{P}}$ and $h \in \overline{\mathcal{H}}_n$ set

$$\widehat{\Delta}_n(h, \mathcal{P}) = \sup_{\eta \in \overline{\mathcal{H}}_n} \sup_{\mathcal{P}' \in \overline{\mathfrak{P}}} \left[\| \widehat{f}_{(h,\mathcal{P}),(\eta,\mathcal{P}')} - \widehat{f}_{\eta,\mathcal{P}'} \|_\infty - \lambda \widehat{A}_n(\eta, \mathcal{P}') \right]_+, \tag{2.2}$$

and let $\widehat{h}$ and $\widehat{\mathcal{P}}$ be defined as follows:

$$\widehat{\Delta}_n(\widehat{h}, \widehat{\mathcal{P}}) + \lambda \widehat{A}_n(\widehat{h}, \widehat{\mathcal{P}}) = \inf_{h \in \overline{\mathcal{H}}_n} \inf_{\mathcal{P} \in \overline{\mathfrak{P}}} \left[\widehat{\Delta}_n(h, \mathcal{P}) + \lambda \widehat{A}_n(h, \mathcal{P}) \right]. \tag{2.3}$$

Our final estimator is $\widehat{f}_{\widehat{h},\widehat{\mathcal{P}}}(x), x \in \mathbb{R}^d$, and let us briefly discuss several issues related to its construction.

Extra parameters $\overline{\mathfrak{P}}$ *and* $\overline{\mathcal{H}}_n$. The necessity to introduce these parameters is dictated only by computational reasons [computation of $\mathbf{f}_n$, $\widehat{\Delta}_n(h, \mathcal{P})$ and minimization led to $\widehat{h}$ and $\widehat{\mathcal{P}}$] and their optimal "theoretical" choice is $\overline{\mathfrak{P}} = \mathfrak{P}$ and $\overline{\mathcal{H}}_n = \mathcal{H}_n$; see discussion after Theorem 1. However, the computational aspects of the choice of $\overline{\mathfrak{P}}$ and $\overline{\mathcal{H}}_n$ are quite different. Typically, $\overline{\mathcal{H}}_n$ can be chosen as an appropriate greed in $\mathcal{H}_n$, for instance diadic one, that is sufficient for proving adaptive properties of the proposed estimator; see Theorem 3.

The choice of $\overline{\mathfrak{P}}$ is much more delicate. The reason we consider $\overline{\mathfrak{P}}$ instead of $\mathfrak{P}$ is explained by the fact that the cardinality of $\mathfrak{P}$ (Bell number) grows as $(d/\ln(d))^d$. Therefore, for large values of d our procedure is not practically feasible in view of the huge amount of comparisons to be done. On the other hand if d is large, the consideration of all partitions is not reasonable itself. Indeed, even theoretically the best attainable trade-off between approximation and stochastic errors corresponds to the effective dimension defined as $d^*(f) = \inf_{\mathcal{P}\in\mathfrak{P}(f)} \sup_{\mathbf{I}\in\mathcal{P}} |\mathbf{I}|$. Of course $d^*(f) \le d$, but if it is proportional, for example, to d, then we will not win much for reasonable sample size. The suitable strategy in the case of large dimension consists of considering only partitions satisfying $\sup_{\mathbf{I}\in\mathcal{P}} |\mathbf{I}| \le d_0$, where d_0 is chosen in accordance with d and the number of observation. In particular one can consider $\overline{\mathfrak{P}}$ containing only 2 elements, namely $\bar{\varnothing}$ and $(\{1\}, \{2\}, \dots, \{d\})$. It corresponds to the hypotheses that we observe vectors with independent components.

Existence and measurability. First, we note that all random fields considered in the paper have continuous trajectories on $\mathcal{H}_n \times \mathbb{R}^d$ in the topology generated by supremum norm. This is guaranteed by Assumption 1. Since $\mathcal{H}_n$ is totally bounded and $\mathbb{R}^d$ can be covered by a countable collection of totally bounded sets, any supremum over $\overline{\mathcal{H}}_n \times \mathbb{R}^d$ of considered random fields will be $X^{(n)}$-measurable. In particular, $\bar{\mathbf{f}}_n$ and

$$\widehat{\Delta}_n(h, \mathcal{P}, \mathcal{P}') := \sup_{\eta\in\overline{\mathcal{H}}_n} [\|\widehat{f}_{(h,\mathcal{P}),(\eta,\mathcal{P}')} - \widehat{f}_{\eta,\mathcal{P}'}\|_\infty - \lambda\widehat{A}_n(\eta, \mathcal{P}')]_+,$$

$$\mathcal{P}, \mathcal{P}' \in \overline{\mathfrak{P}}, h \in \overline{\mathcal{H}}_n.$$

Since $\overline{\mathfrak{P}}$ is finite, we conclude that $\widehat{\Delta}_n(h, \mathcal{P})$ is $X^{(n)}$-measurable for any $\mathcal{P} \in \overline{\mathfrak{P}}$ and any $h \in \mathcal{H}_n$. Assumption 1 implies also that $\widehat{\Delta}_n(\cdot, \mathcal{P})$ and $\widehat{A}_n(\cdot, \mathcal{P})$ are continuous on $\mathcal{H}_n$ for any $\mathcal{P}$. Since $\overline{\mathcal{H}}_n$ is a compact subset of $\mathbb{R}^d$, we conclude that $\widehat{h}(\mathcal{P}) \in \overline{\mathcal{H}}_n$ and $X^{(n)}$-measurable for any $\mathcal{P} \in \overline{\mathfrak{P}}$, Jennrich (1969), where $\widehat{h}(\mathcal{P}) = \inf_{h\in\overline{\mathcal{H}}_n}[\widehat{\Delta}_n(h, \mathcal{P}) + \lambda\widehat{A}_n(h, \mathcal{P})]$. Since $\overline{\mathfrak{P}}$ is finite, we conclude that $(\widehat{h}, \widehat{\mathcal{P}}) \in \overline{\mathcal{H}}_n \times \overline{\mathfrak{P}}$ is $X^{(n)}$-measurable.

Estimation construction. The adaptive and oracle approaches were not developed in the context of multivariate density estimation in the supremum norm, and as a consequence, the selection rule (2.2)–(2.3) leading to the estimator $\widehat{f}_{\widehat{h},\widehat{\mathcal{P}}}(\cdot)$ is new. However, some elements of its construction have been already exploited in previous works.

The idea to use a special estimator based on the convolution operator "$\star$", which led in the considered case to the estimator (2.1), goes back to Lepski and Levit (1999). Further it was steadily used in Goldenshluger and Lepski (2008, 2009, 2011). In particular, if $\overline{\mathfrak{P}} = \{\bar{\varnothing}\}$, that is, the independence hypotheses is not taken into account, our selection rule, based on $\widehat{\Delta}_n(h, \bar{\varnothing})$, is close to the selection rule used in Goldenshluger and Lepski (2011), where the bandwidths selection problem was solved under $\mathbb{L}_s$-loss, $1 \le s < \infty$. In this context, a completely new element of our construction is the adaptation to eventual independence configuration $\boldsymbol{\mathcal{P}} \in \mathfrak{P}(f)$ which is obtained by the introduction of the operation "$\diamond$" on $\mathfrak{P}$ and the criterion $\widehat{\Delta}_n(h, \mathcal{P}), h \in \mathcal{H}_n, \mathcal{P} \in \mathfrak{P}$, based on it. Below we discuss in an informal way the basic facts that led to the selection rule (2.2)–(2.3).

To simplify the presentation of our idea, let us consider the case where $\overline{\mathcal{H}}_n = \{\mathbf{h}\}$ and $\mathbf{h} \in \mathcal{H}_n$ is a given vector. This means that we are not interested in bandwidth selection; for example, the density to be estimated is supposed to belong a given set of smooth functions. If so, the special estimator based on the convolution operator "$\star$" is not needed anymore, and our selection rule can be rewritten on much simpler way. The final estimator is now $\widehat{f}_{\mathbf{h},\widehat{\mathcal{P}}}$ and

$$\widehat{\Delta}_n(\mathbf{h}, \mathcal{P}) = \sup_{\mathcal{P}' \in \overline{\mathfrak{P}}} \big[\|\widehat{f}_{(\mathbf{h},\mathcal{P}\diamond\mathcal{P}')} - \widehat{f}_{\mathbf{h},\mathcal{P}'}\|_\infty - \lambda \widehat{A}_n(\mathbf{h}, \mathcal{P}')\big]_+,$$

$$\widehat{\Delta}_n(\mathbf{h}, \widehat{\mathcal{P}}) + \lambda \widehat{A}_n(\mathbf{h}, \widehat{\mathcal{P}}) = \inf_{\mathcal{P} \in \overline{\mathfrak{P}}} \big[\widehat{\Delta}_n(\mathbf{h}, \mathcal{P}) + \lambda \widehat{A}_n(\mathbf{h}, \mathcal{P})\big].$$

Remember that $\widehat{f}_{\mathbf{h},\mathcal{P}}(x) = \prod_{\mathbf{I}\in\mathcal{P}} \widetilde{f}_{\mathbf{h}_\mathbf{I}}(x_\mathbf{I})$, and $\widetilde{f}_{\mathbf{h}_\mathbf{I}}(x_\mathbf{I})$ is the standard kernel estimator for the marginal density $f_\mathbf{I}(x_\mathbf{I})$.

Set $\xi_{\mathbf{h}_\mathbf{I}}(x_\mathbf{I}) = \widetilde{f}_{\mathbf{h}_\mathbf{I}}(x_\mathbf{I}) - \mathbb{E}_f\{\widetilde{f}_{\mathbf{h}_\mathbf{I}}(x_\mathbf{I})\}$ and note that $\xi_{\mathbf{h}_\mathbf{I}}(x_\mathbf{I}), x_\mathbf{I} \in \mathbb{R}^{|\mathbf{I}|}$, is the sum of i.i.d. bounded and centered random variables and, therefore, is *somehow small*.

If we admit the latter remark we can expect that

$$\begin{aligned}&\|\widehat{f}_{(\mathbf{h},\mathcal{P}\diamond\mathcal{P}')} - \widehat{f}_{\mathbf{h},\mathcal{P}'}\|_\infty \\ &\qquad \approx \bigg\| \prod_{\mathbf{I}\in\mathcal{P}\diamond\mathcal{P}'} \mathbb{E}_f\{\widetilde{f}_{\mathbf{h}_\mathbf{I}}(\cdot)\} - \prod_{\mathbf{I}\in\mathcal{P}'} \mathbb{E}_f\{\widetilde{f}_{\mathbf{h}_\mathbf{I}}(\cdot)\}\bigg\|_\infty + \text{smaller order term.}\end{aligned}$$

The key observations in this context [explaining the introduction of $\widehat{\Delta}_n(\mathbf{h}, \mathcal{P})$] are the following:

$$\prod_{\mathbf{I}\in\mathcal{P}\diamond\mathcal{P}'} \mathbb{E}_f\{\widetilde{f}_{\mathbf{h}_\mathbf{I}}(\cdot)\} \equiv \prod_{\mathbf{I}\in\mathcal{P}'} \mathbb{E}_f\{\widetilde{f}_{\mathbf{h}_\mathbf{I}}(\cdot)\} \qquad \forall \mathcal{P} \in \mathfrak{P}(f), \forall \mathcal{P}' \in \mathfrak{P},$$

and (smaller order term) $< \lambda \widehat{A}_n(\mathbf{h}, \mathcal{P}')$ *with high probability*. Thus *with high probability* $\widehat{\Delta}_n(\mathbf{h}, \mathcal{P}) = 0$ for any $\mathcal{P} \in \mathfrak{P}(f)$.

It allows us to conclude that *with high probability* our procedure select $\widehat{\mathcal{P}} \in \mathfrak{P}(f)$, which, moreover, minimizes $V(\mathbf{h}, \mathcal{P})$. For instance if $(\{1\}, \{2\}, \ldots, \{d\}) =:$

$\mathcal{P}^{(\text{optimal})} \in \mathfrak{P}(f)$, that is, the coordinates of observable vectors are independent random variables, then our procedure selects $\mathcal{P}^{(\text{optimal})}$ *with high probability*. It provides us with the estimator $\widehat{f}_{\mathbf{h},\mathcal{P}^{(\text{optimal})}}$ whose accuracy is dimension free.

2.2. *Main result*. Let $\mathbf{f} > 0$ be a given number, and introduce the following set of densities:

$$\mathbf{F}(\mathbf{f}) = \Big\{ f : \sup_{\mathbf{I} \in \mathcal{I}_d} \| f_{\mathbf{I}} \|_\infty \le \mathbf{f} \Big\}.$$

With any density $f \in \mathbf{F}(\mathbf{f})$, any $h \in (0,1]^d$ and $\mathbf{I} \in \mathcal{I}_d$ associate the quantity

$$b_{h_{\mathbf{I}}} := \left\| \int_{\mathbb{R}^{|\mathbf{I}|}} K_{h_{\mathbf{I}}}(t_{\mathbf{I}} - \cdot)\big[f_{\mathbf{I}}(t_{\mathbf{I}}) - f_{\mathbf{I}}(\cdot)\big]\,\mathrm{d}t_{\mathbf{I}} \right\|_{\mathbf{I},\infty},$$

which can be viewed as the approximation error of $f_{\mathbf{I}}$ measured in the supremum norm.

For any $h \in \mathcal{H}_n$ and $\mathcal{P} \in \mathfrak{P}$ set $B(h,\mathcal{P}) = \sup_{\mathcal{P}' \in \mathfrak{P}} \sup_{\mathbf{I} \in \mathcal{P} \diamond \mathcal{P}'} \|b_{h_{\mathbf{I}}}\|_{\mathbf{I},\infty}$ and introduce the quantity

$$\mathfrak{R}_n(f) = \inf_{h \in \overline{\mathcal{H}}_n} \inf_{\mathcal{P} \in \mathfrak{P}(f) \cap \overline{\mathfrak{P}}} \left(B(h,\mathcal{P}) + \sqrt{\frac{\ln(n)}{nV(h,\mathcal{P})}} \right).$$

THEOREM 1. *Let Assumption 1 be fulfilled. Then for any $q \ge 1$ and any $0 < \mathbf{f} < \infty$ there exist $0 < C_1 < \infty$ and $0 < C_2 < \infty$ such that for any $f \in \mathbf{F}(\mathbf{f})$ and any $n \ge 3$,*

$$\big(\mathbb{E}_f \| \widehat{f}_{\widehat{h},\widehat{\mathcal{P}}} - f \|_\infty^q \big)^{1/q} \le C_1 \mathfrak{R}_n(f) + C_2 n^{-1/2}.$$

The explicit expression of $C_1 = C_1(q,d,\mathbf{K},\mathbf{f})$ and $C_2 = C_2(q,d,\mathbf{K},\mathbf{f})$ can be found in the proof of the theorem.

Let us return to the discussion about extra parameters $\overline{\mathfrak{P}}$ and $\overline{\mathcal{H}}_n$. Their optimal choice is given by $\overline{\mathfrak{P}} = \mathfrak{P}$ and $\overline{\mathcal{H}}_n = \mathcal{H}_n$ since it minimizes obviously the quantity $\mathfrak{R}_n(f)$ for any f. But as it was mentioned above this choice leads to intractable computations of the estimator in the case of large dimension. On the other hand the consideration of $\overline{\mathfrak{P}}$ instead of $\mathfrak{P}$ has a price to pay. It is possible that $\mathfrak{P}(f) \cap \overline{\mathfrak{P}} = \bar{\varnothing}$ although $\mathfrak{P}(f)$ contains the elements besides $\bar{\varnothing}$. However even in this case, where structural hypothesis fails or is not taken into account ($\overline{\mathfrak{P}} = \{\bar{\varnothing}\}$), our estimator solves completely the bandwidths selection problem in multivariate density model under sup-norm loss. Moreover the solution of the considered problem was not known for any $d \ge 2$, and, therefore, our investigations are not restricted by the study of large dimension. In this context, noting that $|\mathfrak{P}| = 2$, $d = 2$, $|\mathfrak{P}| = 5$, $d = 3$, $|\mathfrak{P}| = 12$, $d = 4$ etc, we can conclude that in the case of *small dimension* the optimal choice $\overline{\mathfrak{P}} = \mathfrak{P}$ and $\overline{\mathcal{H}}_n = \mathcal{H}_n$ is always possible. We finish this discussion with the following remark concerning the proof of Theorem 1.

REMARK 1. Our selection rule is based on computation of upper functions for some special type of random processes and the main ingredient of the proof of Theorem 1 is exponential inequality related to them. Corresponding results may have an independent interest and Section 4.1 is devoted to this topic. In particular the function γ_p involved in the construction of our selection rule and which we present below comes from this consideration.

2.3. *Quantity γ_p.* For any $a > 0$, $p \geq 1$ and $s \in \mathbb{N}^*$, introduce

$$\begin{aligned}\gamma_p(s,a) &= 4e\sqrt{2s\tau_p(s,a)\big[a + (3L/2)(a)^{s-1}\big]}\\&\quad + (16e/3)\big(s\big[a + (3L/2)a^{s-1}\big] \vee 8a\big)\tau_p(s,a);\\ \tau_p(s,a) &= s\big(234s\delta_*^{-2} + 6.5p + 5.5\big)\ln(2) + s(2p+3)\\&\quad + \big[108s\delta_*^{-2}\big|\log(a)\big| + 36C_s + 1\big]\big[\ln(3)\big]^{-1}.\end{aligned}$$

Here δ_* is the smallest solution of the equation $8\pi^2\delta(1 + [\ln\delta]^2) = 1$, $C_s = C_s^{(1)} + C_s^{(2)}$ and

$$\begin{aligned}C_s^{(1)} &= s \sup_{\delta>\delta_*} \delta^{-2}\Bigg\{\Bigg[1 + \ln\bigg(\frac{9216(s+1)\delta^2}{[\phi(\delta)]^2}\bigg)\Bigg]_+\\&\quad + 1.5\Bigg[\log_2\bigg\{\bigg(\frac{4608(s+1)\delta^2}{[\phi(\delta)]^2}\bigg)\bigg\}\Bigg]_+\Bigg\};\\ C_s^{(2)} &= s \sup_{\delta>\delta_*} \delta^{-1}\Bigg\{\Bigg[1 + \ln\bigg(\frac{9216(s+1)\delta}{\phi(\delta)}\bigg)\Bigg]_+\\&\quad + 1.5\Bigg[\log_2\bigg\{\bigg(\frac{4608(s+1)\delta}{\phi(\delta)}\bigg)\bigg\}\Bigg]_+\Bigg\},\end{aligned}$$

where $\phi(\delta) = (6/\pi^2)(1 + [\ln\delta]^2)^{-1}$, $\delta > 0$.

3. Adaptive estimation. In this section we illustrate the use of the oracle inequality proved in Theorem 1 for the derivation of adaptive rate optimal density estimators.

We start with the definition of the *anisotropic Nikol'skii class of functions* on $\mathbb{R}^s$, $s \geq 1$, and later on $\mathbf{e}_1, \dots, \mathbf{e}_s$, denotes the canonical basis in $\mathbb{R}^s$.

DEFINITION 1. Let $r = (r_1, \dots, r_s)$, $r_i \in [1, \infty]$, $\alpha = (\alpha_1, \dots, \alpha_s)$, $\alpha_i > 0$, $Q = (Q_1, \dots, Q_s)$, $Q_i > 0$. A function $g : \mathbb{R}^s \to \mathbb{R}$ belongs to the anisotropic Nikol'ski class $\mathbb{N}_{r,s}(\alpha, Q)$ of functions if

$$\|D_i^k g\|_{r_i} \leq Q_i \qquad \forall k = \overline{0, \lfloor\alpha_i\rfloor}, \forall i = \overline{1, s};$$

$$\|D_i^{\lfloor\alpha_i\rfloor} g(\cdot + t\mathbf{e}_i) - D_i^{\lfloor\alpha_i\rfloor} g(\cdot)\|_{r_i} \leq Q_i|t|^{\alpha_i - \lfloor\alpha_i\rfloor} \qquad \forall t \in \mathbb{R}, \forall i = \overline{1, s}.$$

Here $D_i^k f$ denotes the kth order partial derivative of f with respect to the variable t_i, and $\lfloor \alpha_i \rfloor$ is the largest integer strictly less than α_i.

The functional classes $\mathbb{N}_{r,s}(\alpha, Q)$ were considered in approximation theory by Nikol'skii; see, for example, Nikol'skiĭ (1977). Minimax estimation of densities from the class $\mathbb{N}_{r,s}(\alpha, Q)$ was considered in Ibragimov and Khasminskii (1981). We refer also to Kerkyacharian, Lepski and Picard (2001, 2007), where the problem of adaptive estimation over a scale of classes $\mathbb{N}_{r,s}(\alpha, Q)$ was treated for the Gaussian white noise model.

Our goal now is to introduce the scale of functional classes of d-variate probability densities taking into account the independence structure. It implies in particular that we will need to estimate, not only the density itself, but all marginal densities as well. It is easily seen that if $f \in \mathbb{N}_{p,d}(\beta, \mathcal{L})$ and additionally f is compactly supported, then $f_{\mathbf{I}} \in \mathbb{N}_{p_{\mathbf{I}},|\mathbf{I}|}(\beta_{\mathbf{I}}, \overline{\mathcal{L}}_{\mathbf{I}})$ for any $\mathbf{I} \in \mathcal{I}_d$, where $\overline{\mathcal{L}} = \mathrm{c}\mathcal{L}$ and $\mathrm{c} > 0$ is a numerical constant. However if $\operatorname{supp}(f) = \mathbb{R}^d$, the latter assertion is not true in general. The assumption $f \in \mathbb{N}_{p,d}(\beta, \mathcal{L})$ does not even guarantee that $f_{\mathbf{I}}$ is bounded on $\mathbb{R}^{|\mathbf{I}|}$. It explains the introduction of the following anisotropic classes of densities.

Let $p = (p_1, \dots, p_d)$, $p_i \in [1, \infty]$, $\beta = (\beta_1, \dots, \beta_d)$, $\beta_i > 0$, $\mathcal{L} = (\mathcal{L}_1, \dots, \mathcal{L}_d)$, $\mathcal{L}_i > 0$.

DEFINITION 2. A probability density $f : \mathbb{R}^d \to \mathbb{R}_+$ belongs to the class $\overline{\mathbb{N}}_{p,d}(\beta, \mathcal{L})$ if

$$f_{\mathbf{I}} \in \mathbb{N}_{p_{\mathbf{I}},|\mathbf{I}|}(\beta_{\mathbf{I}}, \mathcal{L}_{\mathbf{I}}) \qquad \forall \mathbf{I} \in \mathcal{I}_d.$$

Introduce finally the collection of functional classes taking into account the smoothness of the underlying density and the independence structure simultaneously.

Let $(\beta, p, \mathcal{P}) \in (0, \infty)^d \times [1, \infty]^d \times \mathfrak{P}$ and $\mathcal{L} \in (0, \infty)^d$ be fixed. Introduce

$$\mathbf{N}_{p,d}(\beta, \mathcal{L}, \mathcal{P}) = \left\{ f(x) \in \overline{\mathbb{N}}_{p,d}(\beta, \mathcal{L}) : f(x) = \prod_{\mathbf{I} \in \mathcal{P}} f_{\mathbf{I}}(x_{\mathbf{I}}), \forall x \in \mathbb{R}^d \right\}.$$

For any $(\beta, p, \mathcal{P}) \in (0, \infty)^d \times [1, \infty]^d \times \mathfrak{P}$ define

$$\Upsilon(\beta, p, \mathcal{P}) = \inf_{\mathbf{I} \in \mathcal{P}} \gamma_{\mathbf{I}}(\beta, p), \qquad \gamma_{\mathbf{I}}(\beta, p) = \frac{1 - \sum_{j \in \mathbf{I}} 1/(\beta_j p_j)}{\sum_{j \in \mathbf{I}} 1/\beta_j}.$$

We will see that the quantity $\Upsilon(\beta, p, \mathcal{P})$ can be viewed as an *effective smoothness index* related to independence structure hypothesis and to the estimation under sup-norm loss.

THEOREM 2. *For any* $(\beta, p, \mathcal{P}) \in (0, \infty)^d \times [1, \infty]^d \times \mathfrak{P}$ *such that* $\Upsilon(\beta, p, \mathcal{P}) > 0$ *and any* $\mathcal{L} \in (0, \infty)^d$

$$\liminf_{n\to\infty} \inf_{\widehat{f}_n} \sup_{f \in \mathbf{N}_{p,d}(\beta,\mathcal{L},\mathcal{P})} \left(\mathbb{E}_f^{(n)}[\varphi_n^{-1}(\beta, p, \mathcal{P}) \|\widehat{f}_n - f\|_\infty]^q\right)^{1/q} > 0,$$

$$\varphi_n(\beta, p, \mathcal{P}) = \left(\frac{\ln n}{n}\right)^{\Upsilon/(2\Upsilon+1)},$$

where $\Upsilon = \Upsilon(\beta, p, \mathcal{P})$ *and infimum is taken over all possible estimators.*

Our goal is to prove that the estimation quality provided by $\widehat{f}_{\widehat{h},\widehat{\mathcal{P}}}$ on $\mathbf{N}_{p,d}(\beta, \mathcal{L}, \mathcal{P})$ coincides up to a numerical constant with optimal decay of minimax risk $\varphi_n(\beta, p, \mathcal{P})$ whatever the value of nuisance parameter $\{\beta, p, \mathcal{P}, \mathcal{L}\}$. It means that this estimator is optimally adaptive over the scale of considered functional classes. We would like to emphasize that not only is the couple $(\beta, \mathcal{L})$ unknown, which is typical in frameworks of adaptive estimation, but also the index p of norms where the smoothness is measured. At last, our estimator adapts automatically to unknown independence structure. Note, however, that the range of adaptation with respect to the parameter $\mathcal{P}$ is limited by the set where our procedure is running. It means that, if $\overline{\mathfrak{P}}$ is used in the selection rule, the adaptation is possible only over the collection $\{\mathbf{N}_{\cdot,d}(\cdot, \cdot, \mathcal{P}), \mathcal{P} \in \overline{\mathfrak{P}}\}$. In this context the adaptation over full collection of anisotropic Nikolskii classes is possible only if the selection rule (2.2)–(2.3) runs $\overline{\mathfrak{P}} = \mathfrak{P}$.

Remember that $\bar{\varnothing} = \{1, \dots, d\}$ and, therefore, $\gamma_{\bar{\varnothing}}(\beta, p) = (1 - \sum_{j=1}^d \frac{1}{\beta_j p_j}) \times (\sum_{j=1}^d \frac{1}{\beta_j})^{-1}$. Let $\overline{\mathfrak{P}} \subseteq \mathfrak{P}$ be fixed, and let $\overline{\mathcal{H}}_n$ be the diadic grid in $\mathcal{H}_n$. Let finally $\widehat{f}_{\widehat{h},\widehat{\mathcal{P}}}(\cdot)$ be the estimator obtained by the selection rule (2.2)–(2.3).

THEOREM 3. *Let* $\mathbf{K}$ *satisfy Assumption* 1 *and suppose additionally that for some integer* $\mathfrak{b} \geq 2$,

$$\int_{\mathbb{R}} u^m \mathbf{K}(u)\, du = 0 \qquad \forall m = \overline{2, \mathfrak{b}}. \tag{3.1}$$

Then for any $(\beta, p) \in (0, \mathfrak{b}]^d \times [1, \infty]^d$ *such that* $\gamma_{\bar{\varnothing}}(\beta, p) > 0$, *any* $\mathcal{P} \in \overline{\mathfrak{P}}$ *and any* $\mathcal{L} \in (0, \infty)^d$,

$$\limsup_{n\to\infty} \sup_{f \in \mathbf{N}_{p,d}(\beta,\mathcal{L},\mathcal{P})} \left(\mathbb{E}_f^{(n)}[\varphi_n^{-1}(\beta, p, \mathcal{P}) \|\widehat{f}_{\widehat{h},\widehat{\mathcal{P}}} - f\|_\infty]^q\right)^{1/q} < \infty.$$

Some remarks are in order.

1^0. We want to emphasize that the extra-parameter $\mathfrak{b}$ can be arbitrary but a priory chosen. Note also that the condition (3.1) of the theorem is fulfilled with $m = 1$ as well since $\mathbf{K}$ is symmetric.

2^0. Note that assumption $\gamma_{\bar{\varnothing}}(\beta, p) > 0$ implies obviously $\gamma_{\mathbf{I}}(\beta, p) > 0$ for any $\mathbf{I} \in \mathcal{I}_d$. This together with the definition of the class $\overline{\mathbb{N}}_{p,d}(\beta, \mathcal{L})$ allows us to assert that one can find $\mathbf{f} = \mathbf{f}(\beta, p)$ such that $f \in \overline{\mathbb{N}}_{p,d}(\beta, \mathcal{L})$ implies that $f \in \mathbf{F}(\mathbf{f})$. It makes possible the application of Theorem 1. The latter result follows from the embedding theorem for anisotropic Nikolskii classes which is formulated in the proof of Lemma 4 below.

3^0. As it was recently proven in Goldenshluger and Lepski (2012), Theorem 3(ii), the assumption $\gamma_{\bar{\varnothing}}(\beta, p) > 0$ is necessary for existence of a uniformly consistent estimator of the density f in the supremum norm over anisotropic Nikolskii class $\mathbb{N}_{p,d}(\beta, \mathcal{L})$. Since we consider only $\overline{\mathfrak{P}}$ containing $\bar{\varnothing}$, this means that the estimation of the entire density f is always supposed, the condition $\gamma_{\bar{\varnothing}}(\beta, p) > 0$ is necessary for the application of the selection rule (2.2)–(2.3).

4. Proofs. We start this section with the computation of upper functions for kernel estimation process being one of main tools in the proof of Theorem 1.

4.1. *Upper functions for kernel estimation process.* Let $s \in \mathbb{N}^*$, and let Y_j, $j \geq 1$, be $\mathbb{R}^s$-valued i.i.d. random vectors defined on a complete probability space $(\Omega, \mathfrak{A}, \mathrm{P})$ and having the density $\mathbf{g}$ with respect to the Lebesgue measure. Later on $\mathbb{P}_{\mathbf{g}}^{(n)}$ denotes the law of $Y_1, \ldots, Y_n$, $n \in \mathbb{N}^*$, and $\mathbb{E}_{\mathbf{g}}^{(n)}$ is mathematical expectation with respect to $\mathbb{P}_{\mathbf{g}}^{(n)}$.

Let $\mathbf{M} : \mathbb{R} \to \mathbb{R}$ be a given symmetric function and for any $r \in (0, 1]^s$ set as previously

$$M_r(\cdot) = \prod_{l=1}^{s} r_l^{-1} \mathbf{M}(\cdot / r_l), \qquad V_r = \prod_{l=1}^{s} r_l.$$

Denote also $\mathrm{m}_m = \|\mathbf{M}\|_m$, $m = \{1, \infty\}$. For any $y \in \mathbb{R}^s$ consider the family of random fields

$$\chi_r(y) = n^{-1} \sum_{j=1}^{n} \{ M_r(Y_j - y) - \mathbb{E}_{\mathbf{g}}^{(n)}[M_r(Y_j - y)] \},$$

$$r \in \widetilde{\mathcal{R}}_n(s) := \{ r \in (0, 1]^s : n V_r \geq \ln(n) \}.$$

For any $r \in (0, 1]^s$ set $G(r) = \sup_{y \in \mathbb{R}^s} \int_{\mathbb{R}^s} |M_r(x - y)| \mathbf{g}(x) \, \mathrm{d}x$ and let $\bar{G}(r) = 1 \vee G(r)$.

PROPOSITION 1. *Let $\mathbf{M}$ satisfy Assumption 1. Then for any $n \geq 3$ and any $p \geq 1$,*

$$\mathbb{E}_{\mathbf{g}}^{(n)} \left\{ \sup_{r \in \widetilde{\mathcal{R}}_n(s)} \left[\|\chi_r\|_\infty - \gamma_p(s, \mathrm{m}_\infty) \sqrt{\frac{\bar{G}(r) \ln(n)}{n V_r}} \right] \right\}_+^p$$

$$\leq c_1(p, s) [1 \vee \mathrm{m}_1^s \|\mathbf{g}\|_\infty]^{p/2} n^{-p/2} + c_2(p, s) n^{-p},$$

where $c_1(p, s) = 2^{7p/2+5} 3^{p+5s+4} \Gamma(p + 1) \pi^p(s, \mathrm{m}_\infty)$ and $c_2(p, s) = 2^{p+1} 3^{5s}$.

The function $\pi : \mathbb{N}^* \times \mathbb{R}_+ : \to \mathbb{R}$ is given by

$$\pi(s,a) = (\sqrt{a} \vee a)\big(\sqrt{2es[1+(3L/2)a^{s-2}]} \vee [(2e/3)\big(s[1+(3L/2)a^{s-2}] \vee 8)]\big).$$

In view of trivial inequality,

$$\|\chi_r\|_\infty \le \gamma_p(s,\mathrm{m}_\infty)\sqrt{\frac{\bar{G}(r)\ln(n)}{nV_r}} + \bigg(\|\chi_r\|_\infty - \gamma_p(s,\mathrm{m}_\infty)\sqrt{\frac{\bar{G}(r)\ln(n)}{nV_r}}\bigg)_+,$$

and we come to the following corollary of Proposition 1.

COROLLARY 1. *Let* **M** *satisfy Assumption* 1. *Then for any* $n \ge 3$ *and any* $p \ge 1$,

$$\Big(\mathbb{E}_{\mathbf{g}}^{(n)}\Big\{\sup_{r \in \widetilde{\mathcal{R}}_n(s)} \|\chi_r\|_\infty\Big\}^p\Big)^{1/p}$$
$$\le [1 \vee \mathrm{m}_1^s \|\mathbf{g}\|_\infty]^{1/2}\big[\gamma_p(s,\mathrm{m}_\infty) + \{c_1(p,s)+c_2(p,s)\}^{1/p} n^{-1/2}\big].$$

Consider now the following family of random fields: for any $y \in \mathbb{R}^s$ set

$$\Upsilon_r(y) = n^{-1}\sum_{j=1}^n |M_r(Y_j - y)|, \qquad r \in \widetilde{\mathcal{R}}_n^{(\mathfrak{a})}(s) := \{r \in (0,1]^s : nV_r \ge \mathfrak{a}^{-1}\ln(n)\},$$

where we have put $\mathfrak{a} = [2\gamma_p(s,\mathrm{m}_\infty)]^{-2}$.

PROPOSITION 2. *Let* **M** *satisfy Assumption* 1. *Then for any* $n \ge 3$ *and any* $p \ge 1$,

$$\mathbb{E}_{\mathbf{g}}^{(n)}\Big\{\sup_{r \in \mathcal{R}_n^{(\mathfrak{a})}(s)} [1 \vee \|\Upsilon_r\|_\infty - (3/2)\bar{G}(r)]\Big\}_+^p$$
$$\le c_1(p,s)[1 \vee \mathrm{m}_1^s \|\mathbf{g}\|_\infty]^{p/2} n^{-p/2} + c_2(p,s) n^{-p};$$
$$\mathbb{E}_{\mathbf{g}}^{(n)}\Big\{\sup_{r \in \mathcal{R}_n^{(\mathfrak{a})}(s)} [\bar{G}(r) - 2(1 \vee \|\Upsilon_r(\cdot)\|_\infty)]\Big\}_+^p$$
$$\le c_1'(p,s)[1 \vee \mathrm{m}_1^s \|\mathbf{g}\|_\infty]^{p/2} n^{-p/2} + c_2'(p,s) n^{-p},$$

where $c_1'(p,s) = 2^p c_1(p,s)$ *and* $c_2'(p,s) = 2^{2p+1}3^{5s}$.

4.2. *Proof of Theorem* 1. We start the proof of the theorem with auxiliary results used in the sequel whose proofs are given in Appendix.

4.2.1. *Auxiliary results.* Introduce the following notation. For any $\mathbf{I} \in \mathcal{I}_d$, set

$$s_{h_\mathbf{I}}(\cdot) = \int_{\mathbb{R}^{|\mathbf{I}|}} K_{h_\mathbf{I}}(t_\mathbf{I} - \cdot) f_\mathbf{I}(t_\mathbf{I})\,\mathrm{d}t_\mathbf{I}, \qquad s^*_{h_\mathbf{I},\eta_\mathbf{I}}(\cdot) = \int_{\mathbb{R}^{|\mathbf{I}|}} [K_{h_\mathbf{I}} \star K_{\eta_\mathbf{I}}](t_\mathbf{I} - \cdot) f_\mathbf{I}(t_\mathbf{I})\,\mathrm{d}t_\mathbf{I}.$$

LEMMA 1. *For any $\mathbf{I} \in \mathcal{I}_d$ and any $h, \eta \in (0,1]^{|\mathbf{I}|}$, one has*

$$\|s^*_{h_\mathbf{I},\eta_\mathbf{I}} - s_{\eta_\mathbf{I}}\|_{\mathbf{I},\infty} \le \mathrm{k}_1^d b_{h_\mathbf{I}}.$$

For any $h \in (0,1]^d$ and any $\mathcal{P} \in \mathfrak{P}$, let

$$A_n(h,\mathcal{P}) = \sqrt{\frac{\bar{s}_n \ln(n)}{nV(h,\mathcal{P})}}, \qquad \bar{s}_n = 1 \vee \sup_{h\in\mathcal{H}_n} \sup_{\mathbf{I}\in\mathcal{I}_d} \left\| \int_{\mathbb{R}^{|\mathbf{I}|}} \left|K_{h_\mathbf{I}}(t_\mathbf{I} - \cdot)\right| f_\mathbf{I}(t_\mathbf{I})\,\mathrm{d}t_\mathbf{I} \right\|_{\mathbf{I},\infty}.$$

Put also $\xi_{h_\mathbf{I}}(\cdot) = \widetilde{f}_{h_\mathbf{I}}(\cdot) - s_{h_\mathbf{I}}(\cdot)$, and let

$$\zeta(h,\mathcal{P}) = \sup_{\mathbf{I}\in\mathcal{P}} \|\xi_{h_\mathbf{I}}\|_{\mathbf{I},\infty}, \qquad \zeta_n = \sup_{\eta\in\mathcal{H}_n} \sup_{\mathcal{P}\in\mathfrak{P}} \big[\zeta(\eta,\mathcal{P}) - \Lambda A_n(\eta,\mathcal{P})\big]_+.$$

Set finally,

$$\bar{\mathfrak{f}}_n = \tilde{\mathfrak{f}}_n \mathfrak{f}_n, \qquad \tilde{\mathfrak{f}}_n = d(\bar{\mathbf{f}}_n)^{d^2/4}, \qquad \mathfrak{f}_n = 2d\mathrm{k}_1^d\big[\max\{\bar{\mathbf{f}}_n, \mathrm{k}_1^2\mathbf{f}\}\big]^{d-1}.$$

LEMMA 2. *For any $p \ge 1$ there exist $\mathbf{c}_i(p,d,\mathbf{K},\mathbf{f})$, $\mathbf{i} = 1,2,3,4$, such that for any $n \ge 3$*:

(i) $$\sup_{f\in\mathbf{F}(\mathbf{f})} \big[\mathbb{E}_f^{(n)}(\zeta_n)^{2q}\big]^{1/(2q)} \le \mathbf{c}_1(2q,d,\mathbf{K},\mathbf{f})n^{-1/2};$$

(ii) $$\sup_{f\in\mathbf{F}(\mathbf{f})} \big[\mathbb{E}_f^{(n)}[\bar{s}_n - \bar{\mathbf{f}}_n]_+^{2q}\big]^{1/(2q)} \le \mathbf{c}_2(2q,d,\mathbf{K},\mathbf{f})n^{-1/2};$$

(iii) $$\sup_{f\in\mathbf{F}(\mathbf{f})} \big[\mathbb{E}_f^{(n)}[\bar{\mathbf{f}}_n - 3\bar{s}_n]_+^{2q}\big]^{1/(2q)} \le \mathbf{c}_3(2q,d,\mathbf{K},\mathbf{f})n^{-1/2};$$

(iv) $$\sup_{f\in\mathbf{F}(\mathbf{f})} \big[\mathbb{E}_f^{(n)}(\bar{\mathfrak{f}}_n)^p\big]^{1/p} \le \mathbf{c}_4(p,d,\mathbf{K},\mathbf{f}).$$

The explicit expression of $\mathbf{c}_i(p,d,\mathbf{K},\mathbf{f})$, $\mathbf{i} = 1,2,3,4$ can be found in the proof of the lemma.

4.2.2. *Proof of Theorem 1.* We break the proof on several steps.

1^0. Let $\mathbf{h} \in \overline{\mathcal{H}_n}$ and $\boldsymbol{\mathcal{P}} \in \mathfrak{P}(f) \cap \overline{\mathfrak{P}}$ be fixed. We have, in view of triangle inequality,

$$\begin{aligned}\|\widehat{f}_{\widehat{h},\widehat{\mathcal{P}}} - f\|_\infty \le{}& \|\widehat{f}_{\widehat{h},\widehat{\mathcal{P}}} - \widehat{f}_{(\mathbf{h},\boldsymbol{\mathcal{P}})(\widehat{h},\widehat{\mathcal{P}})}\|_\infty + \|\widehat{f}_{(\mathbf{h},\boldsymbol{\mathcal{P}})(\widehat{h},\widehat{\mathcal{P}})} - \widehat{f}_{\mathbf{h},\boldsymbol{\mathcal{P}}}\|_\infty \\ &+ \|\widehat{f}_{\mathbf{h},\boldsymbol{\mathcal{P}}} - f\|_\infty.\end{aligned} \tag{4.1}$$

We have

$$\|\widehat{f}_{\widehat{h},\widehat{\mathcal{P}}} - \widehat{f}_{(\mathbf{h},\boldsymbol{\mathcal{P}})(\widehat{h},\widehat{\mathcal{P}})}\|_\infty \le \widehat{\Delta}_n(\mathbf{h},\boldsymbol{\mathcal{P}}) + \lambda\widehat{A}_n(\widehat{h},\widehat{\mathcal{P}}). \tag{4.2}$$

Noting that $\widehat{f}_{(\mathbf{h},\boldsymbol{\mathcal{P}})(\widehat{h},\widehat{\mathcal{P}})} \equiv \widehat{f}_{(\widehat{h},\widehat{\mathcal{P}})(\mathbf{h},\boldsymbol{\mathcal{P}})}$, we get

$$\|\widehat{f}_{(\mathbf{h},\boldsymbol{\mathcal{P}})(\widehat{h},\widehat{\mathcal{P}})} - \widehat{f}_{\mathbf{h},\boldsymbol{\mathcal{P}}}\|_\infty \le \widehat{\Delta}_n(\widehat{h},\widehat{\mathcal{P}}) + \lambda\widehat{A}_n(\mathbf{h},\boldsymbol{\mathcal{P}}). \tag{4.3}$$

To obtain (4.2), (4.3) we have also used that $\widehat{\mathcal{P}},\boldsymbol{\mathcal{P}} \in \overline{\mathfrak{P}}$ and $\widehat{h},\mathbf{h} \in \overline{\mathcal{H}}_n$. We get from (4.2) and (4.3),

$$\begin{aligned}
&\|\widehat{f}_{\widehat{h},\widehat{\mathcal{P}}} - \widehat{f}_{(\mathbf{h},\boldsymbol{\mathcal{P}})(\widehat{h},\widehat{\mathcal{P}})}\|_\infty + \|\widehat{f}_{(\mathbf{h},\boldsymbol{\mathcal{P}})(\widehat{h},\widehat{\mathcal{P}})} - \widehat{f}_{\mathbf{h},\boldsymbol{\mathcal{P}}}\|_\infty \\
&\quad \le \big[\widehat{\Delta}_n(\widehat{h},\widehat{\mathcal{P}}) + \lambda\widehat{A}_n(\widehat{h},\widehat{\mathcal{P}})\big] + \big[\widehat{\Delta}_n(\mathbf{h},\boldsymbol{\mathcal{P}}) + \lambda\widehat{A}_n(\mathbf{h},\boldsymbol{\mathcal{P}})\big] \\
&\quad \le 2\big[\widehat{\Delta}_n(\mathbf{h},\boldsymbol{\mathcal{P}}) + \lambda\widehat{A}_n(\mathbf{h},\boldsymbol{\mathcal{P}})\big].
\end{aligned}$$

To get the last inequality we have used the definition of $(\widehat{h},\widehat{\mathcal{P}})$. Thus we obtain from (4.1) that

$$\|\widehat{f}_{\widehat{h},\widehat{\mathcal{P}}} - f\|_\infty \le 2\big[\widehat{\Delta}_n(\mathbf{h},\boldsymbol{\mathcal{P}}) + \lambda\widehat{A}_n(\mathbf{h},\boldsymbol{\mathcal{P}})\big] + \|\widehat{f}_{\mathbf{h},\boldsymbol{\mathcal{P}}} - f\|_\infty. \tag{4.4}$$

We remark that (4.4) remains valid for an arbitrary $\boldsymbol{\mathcal{P}} \in \overline{\mathfrak{P}}$; that is, we did not use that $\boldsymbol{\mathcal{P}} \in \mathfrak{P}(f)$.

2^0. Note that for any $h,\eta \in \mathcal{H}_n$ and any $\mathcal{P}' \in \mathfrak{P}$,

$$\begin{aligned}
&\|\widehat{f}_{(h,\boldsymbol{\mathcal{P}}),(\eta,\mathcal{P}')} - \widehat{f}_{\eta,\mathcal{P}'}\|_\infty \\
&\quad \le |\mathcal{P}'| \sup_{\mathbf{I}'\in\mathcal{P}'}\big((\bar{\mathbf{f}}_n)^{|\mathbf{I}'|(|\mathcal{P}'|-1)}\big) \sup_{\mathbf{I}'\in\mathcal{P}'}\Big\|\prod_{\mathbf{I}\in\boldsymbol{\mathcal{P}}:\,\mathbf{I}\cap\mathbf{I}'\neq\varnothing} \widetilde{f}_{h_{\mathbf{I}\cap\mathbf{I}'},\eta_{\mathbf{I}\cap\mathbf{I}'}} - \widetilde{f}_{\eta_{\mathbf{I}'}}\Big\|_{\mathbf{I}',\infty} \qquad (4.5)\\
&\quad \le d(\bar{\mathbf{f}}_n)^{d^2/4} \sup_{\mathbf{I}'\in\mathcal{P}'}\Big\|\prod_{\mathbf{I}\in\boldsymbol{\mathcal{P}}:\,\mathbf{I}\cap\mathbf{I}'\neq\varnothing} \widetilde{f}_{h_{\mathbf{I}\cap\mathbf{I}'},\eta_{\mathbf{I}\cap\mathbf{I}'}} - \widetilde{f}_{\eta_{\mathbf{I}'}}\Big\|_{\mathbf{I}',\infty}.
\end{aligned}$$

Here $|\mathcal{P}'| = \operatorname{card}(\mathcal{P}')$ and the second inequality follows from $|\mathcal{P}'| \le d + 1 - \sup_{\mathbf{I}'\in\mathcal{P}'}|\mathbf{I}'|$. To get the first bound we have used the trivial inequality: for any $m \in \mathbb{N}^*$ and any $a_j, b_j : \mathcal{X}_j \to \mathbb{R}$, $j = \overline{1,m}$,

$$\begin{aligned}
&\Big\|\prod_{j=1}^m a_j - \prod_{j=1}^m b_j\Big\|_\infty \\
&\quad \le m\Big(\sup_{j=\overline{1,m}}\|a_j - b_j\|_{\mathcal{X}_j,\infty}\Big)\Big[\sup_{j=\overline{1,m}}\max\big(\|a_j\|_{\mathcal{X}_j,\infty}, \|b_j\|_{\mathcal{X}_j,\infty}\big)\Big]^{m-1}, \qquad (4.6)
\end{aligned}$$

where $\|\cdot\|_{\mathcal{X}_j,\infty}$ and $\|\cdot\|_\infty$ denote the supremum norms on $\mathcal{X}_j$ and $\mathcal{X}_1 \times \cdots \times \mathcal{X}_m$, respectively.

Introduce the following notation: for any $h,\eta \in \mathcal{H}_n$ and any $\mathbf{I} \in \mathcal{I}_d$, we set

$$\xi^*_{h_{\mathbf{I}},\eta_{\mathbf{I}}}(\cdot) = \widetilde{f}_{h_{\mathbf{I}},\eta_{\mathbf{I}}}(\cdot) - s^*_{h_{\mathbf{I}},\eta_{\mathbf{I}}}(\cdot).$$

We have in view of (4.6) (here and later the product and the supremum over empty set are assumed equal to one and to zero resp.),

$$\Big\| \prod_{\mathbf{I}\in\mathcal{P}} \widetilde{f}_{h_{\mathbf{I}\cap\mathbf{I}'},\eta_{\mathbf{I}\cap\mathbf{I}'}} - \prod_{\mathbf{I}\in\mathcal{P}} s^*_{h_{\mathbf{I}\cap\mathbf{I}'},\eta_{\mathbf{I}\cap\mathbf{I}'}} \Big\|_{\mathbf{I}',\infty} \tag{4.7}$$
$$\leq d\big[\max\{\bar{\mathbf{f}}_n, \mathrm{k}_1^2\mathbf{f}\}\big]^{d-1} \sup_{\mathbf{I}\in\mathcal{P}} \|\xi^*_{h_{\mathbf{I}\cap\mathbf{I}'},\eta_{\mathbf{I}\cap\mathbf{I}'}}\|_{\mathbf{I}\cap\mathbf{I}',\infty}.$$

We remark that for any $\mathbf{I}\in\mathcal{I}_d$, any $h, \eta \in (0,1]^d$ and any $z_{\mathbf{I}} \in \mathbb{R}^{|\mathbf{I}|}$,

$$\xi^*_{h_{\mathbf{I}},\eta_{\mathbf{I}}}(z_{\mathbf{I}}) = \int_{\mathbb{R}^{|\mathbf{I}|}} K_{\eta_i}(z_{\mathbf{I}} - u_{\mathbf{I}})\xi_{h_{\mathbf{I}}}(u_{\mathbf{I}})\,\mathrm{d}u_{\mathbf{I}}$$

and, therefore,

$$\|\xi^*_{h_{\mathbf{I}},\eta_{\mathbf{I}}}\|_{\mathbf{I},\infty} \leq \mathrm{k}_1^{|\mathbf{I}|}\|\xi_{h_{\mathbf{I}}}\|_{\mathbf{I},\infty} \leq \mathrm{k}_1^d\|\xi_{h_{\mathbf{I}}}\|_{\mathbf{I},\infty},$$

since $\mathrm{k}_1 \geq 1$ in view of Assumption 1. It yields together with (4.7),

$$\Big\| \prod_{\mathbf{I}\in\mathcal{P}} \widetilde{f}_{h_{\mathbf{I}\cap\mathbf{I}'},\eta_{\mathbf{I}\cap\mathbf{I}'}} - \prod_{\mathbf{I}\in\mathcal{P}} s^*_{h_{\mathbf{I}\cap\mathbf{I}'},\eta_{\mathbf{I}\cap\mathbf{I}'}} \Big\|_{\mathbf{I}',\infty} \tag{4.8}$$
$$\leq d\mathrm{k}_1^d\big[\max\{\bar{\mathbf{f}}_n, \mathrm{k}_1^2\mathbf{f}\}\big]^{d-1} \sup_{\mathbf{I}\in\mathcal{P}} \|\xi_{h_{\mathbf{I}\cap\mathbf{I}'}}\|_{\mathbf{I}\cap\mathbf{I}',\infty}.$$

Note also that for any $\eta\in\mathcal{H}_n$ and $\mathbf{I}'\in\mathcal{I}_d$,

$$s_{\eta_{\mathbf{I}'}}(\cdot) = \int_{\mathbb{R}^{\mathbf{I}'}} K_{\eta_{\mathbf{I}'}}(t_{\mathbf{I}'} - \cdot) f_{\mathbf{I}'}(t_{\mathbf{I}'})\,\mathrm{d}t_{\mathbf{I}'} = \int_{\mathbb{R}^{\mathbf{I}'}} K_{\eta_{\mathbf{I}'}}(t_{\mathbf{I}'} - \cdot)\Big[\prod_{\mathbf{I}\in\mathcal{P}} f_{\mathbf{I}\cap\mathbf{I}'}(t_{\mathbf{I}\cap\mathbf{I}'})\Big]\mathrm{d}t_{\mathbf{I}'}$$
$$= \prod_{\mathbf{I}\in\mathcal{P}} s_{\eta_{\mathbf{I}\cap\mathbf{I}'}}(\cdot).$$

Here we have used that $\mathcal{P}\in\mathfrak{P}(f)$. Using once again (4.6), we obtain

$$\Big\| \prod_{\mathbf{I}\in\mathcal{P}} s^*_{h_{\mathbf{I}\cap\mathbf{I}'},\eta_{\mathbf{I}\cap\mathbf{I}'}} - \prod_{\mathbf{I}\in\mathcal{P}} s_{\eta_{\mathbf{I}\cap\mathbf{I}'}} \Big\|_{\mathbf{I}',\infty} \leq d\big[\mathrm{k}_1^2\mathbf{f}\big]^{d-1} \sup_{\mathbf{I}\in\mathcal{P}} \|s^*_{h_{\mathbf{I}\cap\mathbf{I}'},\eta_{\mathbf{I}\cap\mathbf{I}'}} - s_{\eta_{\mathbf{I}\cap\mathbf{I}'}}\|_{\mathbf{I}\cap\mathbf{I}',\infty},$$

and, therefore, in view of Lemma 1,

$$\Big\| \prod_{\mathbf{I}\in\mathcal{P}} s^*_{h_{\mathbf{I}\cap\mathbf{I}'},\eta_{\mathbf{I}\cap\mathbf{I}'}} - s_{\eta_{\mathbf{I}'}} \Big\|_{\mathbf{I}',\infty} \leq d\mathrm{k}_1^d\big[\mathrm{k}_1^2\mathbf{f}\big]^{d-1} \sup_{\mathbf{I}\in\mathcal{P}} \|b_{h_{\mathbf{I}\cap\mathbf{I}'}}\|_{\mathbf{I}\cap\mathbf{I}',\infty}. \tag{4.9}$$

Thus, we obtain from (4.8) and (4.9),

$$\Big\| \prod_{\mathbf{I}\in\mathcal{P}:\,\mathbf{I}\cap\mathbf{I}'\neq\varnothing} \widetilde{f}_{h_{\mathbf{I}\cap\mathbf{I}'},\eta_{\mathbf{I}\cap\mathbf{I}'}} - \widetilde{f}_{\eta_{\mathbf{I}'}} \Big\|_{\mathbf{I}',\infty}$$
$$\leq \mathfrak{f}_n\Big[\sup_{\mathbf{I}\in\mathcal{P}} \|\xi_{h_{\mathbf{I}\cap\mathbf{I}'}}\|_{\mathbf{I}\cap\mathbf{I}',\infty} + \sup_{\mathbf{I}\in\mathcal{P}} \|b_{h_{\mathbf{I}\cap\mathbf{I}'}}\|_{\mathbf{I}\cap\mathbf{I}',\infty}\Big] + \|\xi_{\eta_{\mathbf{I}'}}\|_{\mathbf{I}',\infty},$$

where, remember, $\mathfrak{f}_n = 2d\mathrm{k}_1^d[\max\{\bar{\mathbf{f}}_n, \mathrm{k}_1^2\mathbf{f}\}]^{d-1}$.

Therefore, we get from (4.5) for any $h, \eta \in \mathcal{H}_n$ and $\mathcal{P}' \in \mathfrak{P}$,

$$\begin{aligned}(4.10)\qquad &\|\widehat{f}_{(h,\boldsymbol{\mathcal{P}}),(\eta,\mathcal{P}')} - \widehat{f}_{\eta,\mathcal{P}'}\|_\infty \\ &\quad\le \bar{\mathfrak{f}}_n\Big\{\zeta\big(h, \boldsymbol{\mathcal{P}} \diamond \mathcal{P}'\big) + \sup_{\mathbf{I}\in\boldsymbol{\mathcal{P}}\diamond\mathcal{P}'} \|b_{h_\mathbf{I}}\|_{\mathbf{I},\infty}\Big\} + \tilde{\mathfrak{f}}_n\zeta\big(\eta, \mathcal{P}'\big).\end{aligned}$$

Here, remember, $\tilde{\mathfrak{f}}_n = d(\bar{\mathbf{f}}_n)^{d^2/4}$ and $\bar{\mathfrak{f}}_n = \tilde{\mathfrak{f}}_n\mathfrak{f}_n$. Taking into account that

$$A_n\big(h, \mathcal{P} \diamond \mathcal{P}'\big) \le A_n(h, \mathcal{P}) \wedge A_n\big(h, \mathcal{P}'\big) \qquad \forall h \in \mathcal{H}_n, \forall \mathcal{P}, \mathcal{P}' \in \mathfrak{P},$$

we get from (4.10)

$$\begin{aligned}(4.11)\qquad &\|\widehat{f}_{(h,\boldsymbol{\mathcal{P}}),(\eta,\mathcal{P}')} - \widehat{f}_{\eta,\mathcal{P}'}\|_\infty \\ &\quad\le \bar{\mathfrak{f}}_n\Big\{\Lambda A_n(h, \boldsymbol{\mathcal{P}}) + \sup_{\mathbf{I}\in\boldsymbol{\mathcal{P}}\diamond\mathcal{P}'} \|b_{h_\mathbf{I}}\|_{\mathbf{I},\infty} + \zeta_n\Big\} + \tilde{\mathfrak{f}}_n\zeta\big(\eta, \mathcal{P}'\big).\end{aligned}$$

Remembering that $\lambda = \tilde{\mathfrak{f}}_n\Lambda$, we obtain from (4.11)

$$\begin{aligned}\widehat{\Delta}_n(\mathbf{h}, \boldsymbol{\mathcal{P}}) &\le \bar{\mathfrak{f}}_n\big\{\Lambda A_n(\mathbf{h}, \boldsymbol{\mathcal{P}}) + B(\mathbf{h}, \boldsymbol{\mathcal{P}}) + \zeta_n\big\} \\ &\quad + \tilde{\mathfrak{f}}_n\Big\{\zeta_n + \Lambda \sup_{\eta\in\mathcal{H}_n}\sup_{\mathcal{P}\in\mathfrak{P}}\big[A_n(\eta, \mathcal{P}) - \widehat{A}_n(\eta, \mathcal{P})\big]_+\Big\},\end{aligned}$$

where, remember, $B(h, \mathcal{P}) = \sup_{\mathcal{P}'\in\mathfrak{P}}\sup_{\mathbf{I}\in\mathcal{P}\diamond\mathcal{P}'}\|b_{h_\mathbf{I}}\|_{\mathbf{I},\infty}$. Here we have also used that $\overline{\mathcal{H}}_n \subseteq \mathcal{H}_n$ and $\overline{\mathfrak{P}} \subseteq \mathfrak{P}$. Taking into account that $\bar{\mathfrak{f}}_n \ge \tilde{\mathfrak{f}}_n$, since $\mathfrak{f}_n \ge 1$, we finally get

$$\begin{aligned}(4.12)\qquad \widehat{\Delta}_n(\mathbf{h}, \boldsymbol{\mathcal{P}}) &\le \bar{\mathfrak{f}}_n\Big\{\Lambda A_n(\mathbf{h}, \boldsymbol{\mathcal{P}}) + B(\mathbf{h}, \boldsymbol{\mathcal{P}}) + 2\zeta_n \\ &\quad + \Lambda \sup_{\eta\in\mathcal{H}_n}\sup_{\mathcal{P}\in\mathcal{P}}\big[A_n(\eta, \mathcal{P}) - \widehat{A}_n(\eta, \mathcal{P})\big]_+\Big\}.\end{aligned}$$

Note that the definition of $\mathcal{H}_n$ implies that

$$\big[A_n(\eta, \mathcal{P}) - \widehat{A}_n(\eta, \mathcal{P})\big]_+ \le \mathfrak{a}^*[\sqrt{\bar{s}_n} - \sqrt{\bar{\mathbf{f}}_n}]_+ \le \mathfrak{a}^*[\bar{s}_n - \bar{\mathbf{f}}_n]_+ \qquad \forall \eta \in \mathcal{H}_n, \forall \mathcal{P} \in \mathfrak{P}.$$

To get the last inequality we have also used that $\bar{\mathbf{f}}_n \ge 1$ and $\bar{s}_n \ge 1$ by definition.

Putting $R_n = \mathfrak{a}^*\Lambda[\bar{s}_n - \bar{\mathbf{f}}_n]_+$ we obtain in view of (4.12),

$$(4.13)\qquad \widehat{\Delta}_n(\mathbf{h}, \boldsymbol{\mathcal{P}}) \le \bar{\mathfrak{f}}_n\big\{\Lambda A_n(\mathbf{h}, \boldsymbol{\mathcal{P}}) + B(\mathbf{h}, \boldsymbol{\mathcal{P}}) + 2\zeta_n + R_n\big\}.$$

Note also that the definition of $\mathcal{H}_n$ implies that

$$\big[\widehat{A}_n(\eta, \mathcal{P}) - \sqrt{3}A_n(\eta, \mathcal{P})\big]_+ \le \mathfrak{a}^*[\sqrt{\bar{\mathbf{f}}_n} - \sqrt{3\bar{s}_n}]_+ \le \mathfrak{a}^*[\bar{\mathbf{f}}_n - 3\bar{s}_n]_+ \qquad \forall \eta \in \mathcal{H}_n, \forall \mathcal{P} \in \mathfrak{P}.$$

Thus, denoting $\mathcal{R}_n = \mathfrak{a}^* \Lambda[\bar{\mathbf{f}}_n - 3\bar{s}_n]_+$, we obtain using (4.13),

$$\widehat{\Delta}_n(\mathbf{h}, \boldsymbol{\mathcal{P}}) + \lambda \widehat{A}_n(\mathbf{h}, \boldsymbol{\mathcal{P}}) \le \bar{\mathfrak{f}}_n\{3\Lambda A_n(\mathbf{h}, \boldsymbol{\mathcal{P}}) + B(\mathbf{h}, \boldsymbol{\mathcal{P}}) + 2\zeta_n + R_n + \mathcal{R}_n\}, \tag{4.14}$$

where we have used also $\sqrt{3} < 2$.

3^0. Note that in view of $\boldsymbol{\mathcal{P}} \in \mathfrak{P}(f)$ and (4.6),

$$\begin{aligned} \|\widehat{f}_{\mathbf{h},\boldsymbol{\mathcal{P}}} - f\|_\infty &= \Big\| \prod_{\mathbf{I}\in\boldsymbol{\mathcal{P}}} \widetilde{f}_{\mathbf{h_I}}(x_\mathbf{I}) - \prod_{\mathbf{I}\in\boldsymbol{\mathcal{P}}} f_\mathbf{I}(x_\mathbf{I}) \Big\|_\infty \\ &\le d\big[\max\{\bar{\mathbf{f}}_n, \mathrm{k}_1^2\mathbf{f}\}\big]^{d-1} \sup_{\mathbf{I}\in\boldsymbol{\mathcal{P}}} \|\widetilde{f}_{\mathbf{h_I}}(x_\mathbf{I}) - f_\mathbf{I}(x_\mathbf{I})\|_{\mathbf{I},\infty} \\ &\le d\big[\max\{\bar{\mathbf{f}}_n, \mathrm{k}_1^2\mathbf{f}\}\big]^{d-1}\big[B(\mathbf{h}, \boldsymbol{\mathcal{P}}) + \zeta(\mathbf{h}, \boldsymbol{\mathcal{P}})\big] \\ &\le \bar{\mathfrak{f}}_n\big[B(\mathbf{h}, \boldsymbol{\mathcal{P}}) + \Lambda A_n(\mathbf{h}, \boldsymbol{\mathcal{P}}) + \zeta_n\big]. \end{aligned} \tag{4.15}$$

Here we have also used that $\mathcal{P} \equiv \mathcal{P} \diamond \mathcal{P}$. We obtain from (4.4), (4.14) and (4.15),

$$\|\widehat{f}_{\widehat{h},\widehat{\mathcal{P}}} - f\|_\infty \le \bar{\mathfrak{f}}_n\big[3B(\mathbf{h}, \boldsymbol{\mathcal{P}}) + 7\Lambda A_n(\mathbf{h}, \boldsymbol{\mathcal{P}}) + 5\zeta_n + 2R_n + 2\mathcal{R}_n\big],$$

and, therefore, for any $\mathbf{h} \in \overline{\mathcal{H}}_n$, $\boldsymbol{\mathcal{P}} \in \mathfrak{P}(f) \cap \overline{\mathfrak{P}}$ and $q \ge 1$

$$\begin{aligned} \big(\mathbb{E}_f^{(n)} \|\widehat{f}_{\widehat{h},\widehat{\mathcal{P}}} - f\|_\infty\big)^{1/q} &\le E_q\big[3B(\mathbf{h}, \boldsymbol{\mathcal{P}}) + 7\Lambda A_n(\mathbf{h}, \boldsymbol{\mathcal{P}})\big] \\ &\quad + E_{2q}\big[5y_{1,n} + 2\Lambda\mathfrak{a}^*(y_{2,n} + y_{3,n})\big], \end{aligned} \tag{4.16}$$

where we have put for $p \ge 1$,

$$E_p = \big[\mathbb{E}_f^{(n)}(\bar{\mathfrak{f}}_n)^p\big]^{1/p}, \qquad y_{1,n} = \big[\mathbb{E}_f^{(n)}(\zeta_n)^{2q}\big]^{1/(2q)},$$
$$y_{2,n} = \big[\mathbb{E}_f^{(n)}[\bar{s}_n - \bar{\mathbf{f}}_n]_+^{2q}\big]^{1/(2q)}, \qquad y_{3,n} = \big[\mathbb{E}_f^{(n)}[\bar{\mathbf{f}}_n - 3\bar{s}_n]_+^{2q}\big]^{1/(2q)}.$$

Taking into account that the left-hand side of (4.16) is independent of the choice $\mathbf{h}$ and $\boldsymbol{\mathcal{P}}$ and that the quantity $\bar{s}_n \le 1 \vee [\mathrm{k}_1\mathbf{f}]$, we get

$$\begin{aligned} &\big(\mathbb{E}_f^{(n)} \|\widehat{f}_{\widehat{h},\widehat{\mathcal{P}}} - f\|_\infty\big)^{1/q} \\ &\quad \le 7\Lambda E_q\Big(\inf_{h\in\overline{\mathcal{H}}_n} \inf_{\mathcal{P}\in\mathfrak{P}(f)\cap\overline{\mathfrak{P}}} \big[B(\mathbf{h}, \boldsymbol{\mathcal{P}}) + A_n(\mathbf{h}, \boldsymbol{\mathcal{P}})\big]\Big) \\ &\qquad + E_{2q}\big[5y_{1,n} + 2\Lambda\mathfrak{a}^*(y_{2,n} + y_{3,n})\big] \\ &\quad = \mathbf{C}_1(q, d, \mathbf{K}, \mathbf{f})\mathfrak{R}_n(f) + E_{2q}\big[5y_{1,n} + 2\Lambda\mathfrak{a}^*(y_{2,n} + y_{3,n})\big], \end{aligned}$$

where we have put $\mathbf{C}_1(q, d, \mathbf{K}, \mathbf{f}) = 7\Lambda E_q\sqrt{1 \vee [\mathrm{k}_1\mathbf{f}]}$.

This inequality together with bounds found in Lemma 2 lead to the assertion of the theorem.

4.3. *Proof of Theorem* 2. The proof of Theorem 2 is relatively standard and based on the general result established in Kerkyacharian, Lepski and Picard (2007), Proposition 7. For convenience we formulate this result, not in full generality, but its version reduced to the considered problem. Let $(\beta, p, \mathcal{P}) \in (0,\infty)^d \times [1,\infty]^d \times \mathfrak{P}$ such that $\Upsilon(\beta, p, \mathcal{P}) > 0$ and $\mathcal{L} \in (0,\infty)^d$ be fixed.

LEMMA 3. *Assume that there exist $f_0 \in \mathbf{N}_{p,d}(\beta, \mathcal{L}, \mathcal{P})$, $\rho_n > 0, n \in \mathbb{N}^*$ and a finite set $\mathbf{J}_n$ such that for any sufficiently large $n \in \mathbb{N}^*$, one can find $\{f^{(\mathbf{j})}, \mathbf{j} \in \mathbf{J}_n\} \subset \mathbf{N}_{p,d}(\beta, \mathcal{L}, \mathcal{P})$ satisfying*

$$\|f^{(\mathbf{j})} - f_0\|_\infty = \rho_n \qquad \forall \mathbf{j} \in \mathbf{J}_n; \tag{4.17}$$

$$\limsup_{n\to\infty} \mathbb{E}_{f_0}^{(n)} \left[\frac{1}{|\mathbf{J}_n|} \sum_{\mathbf{j}\in\mathbf{J}_n} \frac{\mathrm{d}\mathbb{P}_{f^{(\mathbf{j})}}^{(n)}}{\mathrm{d}\mathbb{P}_{f_0}^{(n)}} (X^{(n)}) - 1 \right]^2 =: \mathbf{C} < \infty. \tag{4.18}$$

Then for $r \geq 1$,

$$\liminf_{n\to\infty} \inf_{\tilde f} \sup_{f\in\mathbf{N}_{p,d}(\beta,\mathcal{L},\mathcal{P})} \rho_n^{-1} \big(\mathbb{E}_f^{(n)} \|\tilde f - f\|_\infty^r\big)^{1/r} \geq 2^{-1}\big[1 - \sqrt{\mathbf{C}/(\mathbf{C}+4)}\big],$$

where infimum is taken over all possible estimators.

PROOF OF THE THEOREM. Set $\mathcal{N}(x) = \prod_{i=1}^d ([2\pi]^{-1/2} \exp -\{x_i^2/2\})$, and let $f_0(x) = \sigma^{-d}\mathcal{N}(x/\sigma)$, where $\sigma > 0$ is chosen in such a way that

$$f_0 \in \overline{\mathbb{N}}_{p,d}(\beta, \mathcal{L}/2). \tag{4.19}$$

The product structure of f_0 together with (4.19) allows us to assert that $f_0 \in \mathbf{N}_{p,d}(\beta, \mathcal{L}/2, \mathcal{P})$ for any $\mathcal{P} \in \mathfrak{P}$. Let $\mathbf{I}^* \in \{1, \dots, d\}$ be defined from the relation

$$\Upsilon(\beta, p, \mathcal{P}) := \inf_{\mathbf{I}\in\mathcal{P}} \gamma_{\mathbf{I}}(\beta, p) = \gamma_{\mathbf{I}^*}(\beta, p),$$

and for notational convenience, the elements of $\mathbf{I}^*$ will be denoted by $i_1, \dots, i_m$ and $m = |\mathbf{I}^*|$.

Let $g : \mathbb{R} \to \mathbb{R}$ be compactly supported on $(-1/2, 1/2)$ function, satisfying $g \in \bigcap_{i\in\mathbf{I}^*} N_{p_i,1}(\beta_i, 1/2)$, and such that $\int g = 0$. Suppose also that $|g(0)| = \|g\|_\infty$.

Let $A_n \to 0$ and $\delta_{l,n} \to 0$, $l = \overline{1,m}$, $n \to \infty$, be sequences whose choice will be done later and set $\mathbf{J}_n := [1, \dots, M_{1,n}] \times \cdots \times [1, \dots, M_{m,n}] \subset \mathbb{N}^m$, where $M_{l,n} = \lfloor \delta_{l,n}^{-1/2} \rfloor$, $l = \overline{1,m}$.

For any $\mathbf{j} = (j_1, \dots, j_m) \in \mathbf{J}_n$ define $G_{\mathbf{j}}(x_{\mathbf{I}}) = A_n \prod_{l=1}^m g(\delta_{i,n}^{-1}[x_{i_l} - x_{i_l}^{(\mathbf{j})}])$. Here for any $\mathbf{j} \in \mathbf{J}_n$, we put $x_{i_l}^{(\mathbf{j})} = j_l \delta_{l,n}$. The choice of g implies

$$G_{\mathbf{j}} G_{\mathbf{j}'} \equiv 0 \qquad \forall \mathbf{j}, \mathbf{j}' \in \mathbf{J}_n, \mathbf{j} \neq \mathbf{j}'. \tag{4.20}$$

Note also that the system of equations

$$A_n \delta_{k,n}^{-\beta_{i_k}} \left(\prod_{l=1}^m \delta_{l,n} \right)^{1/p_{i_k}} = \frac{\mathcal{L}_{i_k}}{c_k}, \qquad k = \overline{1,m} \tag{4.21}$$

implies that $G_{\mathbf{j}} \in N_{p_{\mathbf{I}},d}(\beta_{\mathbf{I}}, \mathcal{L}_{\mathbf{I}}/2)$ for any $\mathbf{j} \in \mathbf{J}_n$. Here we have denoted $c_k = (\|g\|_{p_{i_k}})^{m-1}$.

Introduce the family of functions $\{f^{(\mathbf{j})}, \mathbf{j} \in \mathbf{J}_n\}$ as follows:

$$f^{(\mathbf{j})}(x) = \prod_{i \notin \mathbf{I}^*}^{d} ([2\pi\sigma^2]^{-1/2} \exp -\{x_i^2/2\sigma^2\}) \times \left(\prod_{i \in \mathbf{I}^*}^{d} [2\pi\sigma^2]^{-1/2} \exp -\{x_i^2/2\sigma^2\} + G_{\mathbf{j}}(x_{\mathbf{I}})\right).$$

First we remark that $A_n \to 0, n \to \infty$, implies that $f^{(\mathbf{j})} > 0$ for all sufficiently large n. Next, the assumption $\int g = 0$ implies that $\int f^{(\mathbf{j})} = 1$. Thus, $f^{(\mathbf{j})}$ is a probability density for any $\mathbf{j} \in \mathbf{J}_n$ for all sufficiently large n. At last the choice of f_0 together with (4.21) allows us to assert that $f^{(\mathbf{j})} \in \mathbf{N}_{p,d}(\beta, \mathcal{L}, \mathcal{P})$ for any $\mathbf{j} \in \mathbf{J}_n$.

Thus, we conclude that Lemma 3 is applicable to the family $\{f^{(\mathbf{j})}, \mathbf{j} \in \mathbf{J}_n\}$. We remark also that

$$\|f^{(\mathbf{j})} - f_0\|_\infty = c_1^* A_n \qquad \forall \mathbf{j} \in \mathbf{J}_n, \tag{4.22}$$

where we have put $c_1^* = |g(0)|^m (2\pi\sigma^2)^{(m-d)/2}$. Here we have also used that $|g(0)| = \|g\|_\infty$. We conclude that assumption (4.17) is fulfilled with $\rho_n = c_1^* A_n$.

Let us now proceed with the verification of condition (4.18) of Lemma 3. Note first that

$$\frac{\mathrm{d}\mathbb{P}^{(n)}_{f^{(\mathbf{j})}}}{\mathrm{d}\mathbb{P}^{(n)}_{f_0}}(X^{(n)}) = \prod_{k=1}^{n} \frac{f^{(\mathbf{j})}(X_k)}{f_0(X_k)}$$

and, therefore,

$$\begin{aligned} &\left[\frac{1}{|\mathbf{J}_n|} \sum_{\mathbf{j} \in \mathbf{J}_n} \frac{\mathrm{d}\mathbb{P}^{(n)}_{f^{(\mathbf{j})}}}{\mathrm{d}\mathbb{P}^{(n)}_{f_0}}(X^{(n)})\right]^2 \\ &\quad = \frac{1}{|\mathbf{J}_n|^2} \left\{ \sum_{\mathbf{j} \in \mathbf{J}_n} \prod_{k=1}^{n} \left[\frac{f^{(\mathbf{j})}(X_k)}{f_0(X_k)}\right]^2 + \sum_{\substack{\mathbf{j},\mathbf{j}' \in \mathbf{J}_n: \\ \mathbf{j} \neq \mathbf{j}'}} \prod_{k=1}^{n} \frac{f^{(\mathbf{j})}(X_k) f^{(\mathbf{j}')}(X_k)}{f_0^2(X_k)} \right\}. \end{aligned} \tag{4.23}$$

Since $X_k, k = \overline{1,n}$ are i.i.d. random vectors, we have for any $\mathbf{j} \neq \mathbf{j}'$,

$$\begin{aligned} &\mathbb{E}^{(n)}_{f_0} \left\{ \prod_{k=1}^{n} \frac{f^{(\mathbf{j})}(X_k) f^{(\mathbf{j}')}(X_k)}{f_0^2(X_k)} \right\} \\ &\quad = \left\{ \int_{\mathbb{R}^{|\mathbf{I}^*|}} \left[1 + \frac{G_{\mathbf{j}}(x_{\mathbf{I}^*})}{f_{\mathbf{I}^*,0}(x_{\mathbf{I}^*})}\right]\left[1 + \frac{G_{\mathbf{j}'}(x_{\mathbf{I}^*})}{f_{\mathbf{I}^*,0}(x_{\mathbf{I}^*})}\right] f_{\mathbf{I}^*,0}(x_{\mathbf{I}^*})\, \mathrm{d}x_{\mathbf{I}^*} \right\}^n = 1. \end{aligned}$$

To get the last equality we have used (4.20) and the fact that $\int_{\mathbb{R}^{|\mathbf{I}^*|}} G_{\mathbf{j}}(x_{\mathbf{I}^*})\,\mathrm{d}x_{\mathbf{I}^*} = 0$ since $\int g = 0$.

The latter result together with (4.23) yields

$$\mathcal{E}_n := \mathbb{E}_{f_0}^{(n)}\biggl[\frac{1}{|\mathbf{J}_n|}\sum_{\mathbf{j}\in\mathbf{J}_n}\frac{\mathrm{d}\mathbb{P}^{(n)}_{f^{(\mathbf{j})}}}{\mathrm{d}\mathbb{P}^{(n)}_{f_0}}(X^{(n)}) - 1\biggr]^2$$

$$(4.24)\qquad = \frac{1}{|\mathbf{J}_n|^2}\sum_{\mathbf{j}\in\mathbf{J}_n}\biggl\{\int_{\mathbb{R}^{|\mathbf{I}^*|}}\biggl[1 + \frac{G_{\mathbf{j}}(x_{\mathbf{I}^*})}{f_{\mathbf{I}^*,0}(x_{\mathbf{I}^*})}\biggr]^2 f_{\mathbf{I}^*,0}(x_{\mathbf{I}^*})\,\mathrm{d}x_{\mathbf{I}^*}\biggr\}^n - |\mathbf{J}_n|^{-1}$$

$$= \frac{1}{|\mathbf{J}_n|^2}\sum_{\mathbf{j}\in\mathbf{J}_n}\biggl\{1 + \int_{\mathbb{R}^m}\biggl[\frac{G^2_{\mathbf{j}}(y)}{f_{\mathbf{I}^*,0}(y)}\biggr]\mathrm{d}y\biggr\}^n - |\mathbf{J}_n|^{-1}.$$

Since $G_{\mathbf{j}}(y) = 0$ for any $y \notin [0, \sqrt{\delta_{1,n}}] \times \cdot \times [0, \sqrt{\delta_{m,n}}] =: \mathcal{Y}_n$, we have for all n large enough $\inf_{y\in\mathcal{Y}_n} f_{\mathbf{I}^*,0}(y) \geq 2^{-1}(2\pi\sigma^2)^{-m}$. It yields together with (4.23), putting $c_2^* = 2(2\pi\sigma^2)^m \|g\|_2^{2m}$,

$$\mathcal{E}_n \leq |\mathbf{J}_n|^{-1}\Biggl(1 + c_2^* A_n^2 \prod_{l=1}^m \delta_{l,n}\Biggr)^n.$$

If we choose A_n and $\delta_{l,n}, l = \overline{1,m}$ satisfying

$$(4.25)\qquad c_2^* n A_n^2 \prod_{l=1}^m \delta_{l,n} \leq (1/4)\ln\Biggl(\prod_{l=1}^m \delta_{l,n}^{-1}\Biggr) \leq \ln\bigl(|\mathbf{J}_n|\bigr)$$

for all $n \geq 1$ large enough, then $\mathcal{E}_n \leq 1$, and therefore, condition (4.18) is fulfilled with $\mathbf{C} = 1$.

Thus, we have to choose A_n and $\delta_{l,n}, l = \overline{1,m}$ satisfying (4.21) and (4.25). Let $t > 0$ be the number whose choice will be done later. Consider instead (4.25) the equation

$$(4.26)\qquad n A_n^2 \prod_{l=1}^m \delta_{l,n} = t^2 \ln(n),$$

and solve (4.21) and (4.26). Straightforward computations yield

$$A_n = R(\varepsilon t)^{(1-\sum_{l=1}^m 1/(\beta_{i_l} p_{i_l}))/(1-\sum_{l=1}^m (1/p_{i_l} - 1/2)1/\beta_{i_l})},$$

$$\delta_{l,n} = A_n^{1/\beta_{i_l} - 2/(\beta_{i_l} p_{i_l})}(t\varepsilon)^{2/(\beta_{i_l} p_{i_l})}(c_l/\mathcal{L}_l)^{1/\beta_{i_l}},$$

where we have put $R = (\prod_{l=1}^m (c_l/\mathcal{L}_l)^{1/(2\beta_{i_l})})^{1/(1-\sum_{l=1}^m (1/p_{i_l} - 1/2)(1/\beta_{i_l}))}$. Moreover we have in view of (4.26)

$$\Biggl(\prod_{l=1}^m \delta_{l,n}\Biggr)^{-1/2} = R(\varepsilon t)^{-a}, \qquad a = \frac{\sum_{l=1}^m 1/\beta_{i_l}}{1 - \sum_{l=1}^m (1/p_{i_l} - 1/2)(1/\beta_{i_l})}$$

and, therefore, $(1/4)\ln(\prod_{l=1}^m \delta_{l,n}^{-1}) \asymp (a/2)\ln(n)$, $n \to \infty$. Hence, choosing t as an arbitrary number satisfying $t^2 < (2c_2^*)^{-1}a$ we guarantee that (4.26) implies (4.25) for all n large enough.

Thus we conclude that Lemma 3 is applicable with

$$\rho_n = c_1^* A_n = c_1^* R\left(\frac{t\ln(n)}{n}\right)^{(1-\sum_{l=1}^m 1/(\beta_{i_l} p_{i_l}))/(2(1-\sum_{l=1}^m [1/p_{i_l}-1/2](1/\beta_{i_l})))}.$$

It remains to note that the definition of I^* implies that $\Upsilon(\beta, p, \mathcal{P}) = \frac{1-\sum_{l=1}^m 1/(\beta_{i_l} p_{i_l})}{\sum_{l=1}^m 1/\beta_{i_l}}$. We remark that

$$\frac{\Upsilon(\beta, p, \mathcal{P})}{2\Upsilon(\beta, p, \mathcal{P})+1} = \frac{1-\sum_{l=1}^m 1/(\beta_{i_l} p_{i_l})}{2(1-\sum_{l=1}^m [1/p_{i_l}-1/2](1/\beta_{i_l}))},$$

and the assertion of the theorem follows. □

4.4. *Proof of Theorem* 3. The proof of the theorem is based on the application of Theorem 1 and on Lemma 4 below that allows us to bound from above the quantity $B(h, \mathcal{P})$. The assertion of the lemma, whose proof is postponed to the Appendix, is based on the embedding theorem for anisotropic Nikolskii classes. For any function $g: \mathbb{R}^s \to \mathbb{R}$ and any $\eta \in (0,\infty)^s$, set

$$\mathcal{B}_{\eta,g}(z) = \int_{\mathbb{R}^s} K_\eta(t-z)g(t)\,\mathrm{d}t - g(z), \qquad z \in \mathbb{R}^s.$$

LEMMA 4. *Let* $\mathbf{K}$ *satisfy Assumption* 1 *and* (3.1). *Let* $(\alpha, r) \in (0, \mathfrak{b}]^s \times [1,\infty]^s$ *be such that* $\varkappa = 1 - \sum_{l=1}^s (\alpha_l r_l)^{-1} > 0$, *and let* $Q \in (0,\infty)^s$. *Then there exists* $\mathrm{c} = \mathrm{c}(s, r, \mathfrak{b}) > 0$ *such that*

$$\sup_{g \in \mathbb{N}_{r,s}(\alpha, Q)} \|\mathcal{B}_{\eta,g}\|_\infty \le \mathrm{c}\mathrm{k}_1^s \sum_{i=1}^s Q_i \eta_i^{\boldsymbol{\alpha}_i}, \qquad \forall \eta \in (0,\infty)^s.$$

Here $\boldsymbol{\alpha} = (\boldsymbol{\alpha}_1, \dots, \boldsymbol{\alpha}_s)$, $\boldsymbol{\alpha}_i = \varkappa\alpha_i\varkappa_i^{-1}$ *and* $\varkappa_i = 1 - \sum_{l=1}^s (r_l^{-1} - r_i^{-1})\alpha_l^{-1}$.

PROOF OF THEOREM 3. Let $(\beta, p, \mathcal{P}) \in (0, \mathfrak{b}]^d \times [1,\infty]^d \times \overline{\mathfrak{P}}$ and $\mathcal{L} \in (0,\infty)^d$ be fixed. For any $\mathbf{I} \in \mathcal{I}_d$ and any $\mathbf{i} \in \mathbf{I}$ define

$$\boldsymbol{\beta}_{\mathbf{i}}(\mathbf{I}) = \tau(\mathbf{I})\beta_{\mathbf{i}}\tau_{\mathbf{i}}^{-1}(\mathbf{I}), \qquad \tau(\mathbf{I}) = 1 - \sum_{l\in\mathbf{I}} (\beta_l p_l)^{-1},$$

$$\tau_{\mathbf{i}}(\mathbf{I}) = 1 - \sum_{l\in\mathbf{I}} (p_l^{-1} - p_{\mathbf{i}}^{-1})\beta_l^{-1},$$

and note that the condition $\gamma_{\bar{\varnothing}}(\beta, p) > 0$ implies that $\tau(\mathbf{I}) > 0$ for any $\mathbf{I} \in \mathcal{I}_d$.

Let us first prove the following simple fact. Denote $\mathcal{C}_{\mathbf{i}}(\mathbf{I}) = \{\mathbf{J} \subseteq \mathbf{I} : \mathbf{i} \in \mathbf{J}\}$, $\mathbf{i} \in \mathbf{I}$. Then

$$\boldsymbol{\beta}_{\mathbf{i}}(\mathbf{I}) = \inf_{\mathbf{J}\in\mathcal{C}_{\mathbf{i}}(\mathbf{I})} \boldsymbol{\beta}_{\mathbf{i}}(\mathbf{J}) \qquad \forall \mathbf{i} \in \mathbf{I}. \tag{4.27}$$

Indeed, we remark that $\tau_{\mathbf{i}}(\mathbf{J}) = 1 - \sum_{l\in\mathbf{J}}(p_l^{-1} - p_{\mathbf{i}}^{-1})\beta_l^{-1} = \tau(\mathbf{J}) + p_{\mathbf{i}}^{-1}\sum_{l\in\mathbf{J}}\beta_l^{-1}$ and, therefore,

$$\boldsymbol{\beta}_{\mathbf{i}}(\mathbf{J}) = \frac{\beta_{\mathbf{i}}\tau(\mathbf{J})}{\tau(\mathbf{J}) + p_{\mathbf{i}}^{-1}\beta^{-1}(\mathbf{J})}, \qquad \beta^{-1}(\mathbf{J}) = \sum_{l\in\mathbf{J}}\beta_l^{-1}.$$

We obviously have $\tau(\mathbf{J}) \geq \tau(\mathbf{I})$ and $\beta^{-1}(\mathbf{J}) \leq \beta^{-1}(\mathbf{I})$ for any $\mathbf{J} \subseteq \mathbf{I}$. It remains to note that $x \mapsto x/(x+a)$ is increasing on $\mathbb{R}_+$ for any $a > 0$, and (4.27) follows.

Let $\mathcal{P}' \in \mathfrak{P}$ be an arbitrary partition. Since $f \in \overline{\mathbb{N}}_{p,d}(\beta, \mathcal{L})$ we have $f_{\mathbf{J}} \in \mathbb{N}_{p_{\mathbf{J}},|\mathbf{J}|}(\beta_{\mathbf{J}}, \mathcal{L}_{\mathbf{J}})$, and therefore, in view of Lemma 4 we have for any $h \in (0,1]^d$ and $\mathbf{J} \in \mathcal{P} \diamond \mathcal{P}'$,

$$b_{h_{\mathbf{J}}} \leq \mathrm{c}(|\mathbf{J}|, p_{\mathbf{J}}, \mathfrak{b})\mathrm{k}_1^{|\mathbf{J}|}\sum_{\mathbf{i}\in\mathbf{J}}\mathcal{L}_{\mathbf{i}}h_{\mathbf{i}}^{\boldsymbol{\beta}_{\mathbf{i}}(\mathbf{J})} \leq \mathrm{c}_1\sum_{\mathbf{i}\in\mathbf{I}}\mathcal{L}_{\mathbf{i}}h_{\mathbf{i}}^{\boldsymbol{\beta}_{\mathbf{i}}(\mathbf{I})}.$$

To get the last inequality we use (4.27), $h \in (0,1]^d$ and we have put $\mathrm{c}_1 = \mathrm{k}_1^d \sup_{\mathbf{J}\in\mathcal{I}_d}\mathrm{c}(|\mathbf{J}|, p_{\mathbf{J}}, \mathfrak{b})\mathrm{k}_1^{|\mathbf{J}|}$.

Noting that the right-hand side of the latter inequality is independent on $\mathbf{J}$, we obtain

$$B(h,\mathcal{P}) \leq \mathrm{c}_1\sup_{\mathbf{I}\in\mathcal{P}}\sum_{\mathbf{i}\in\mathbf{I}}\mathcal{L}_{\mathbf{i}}h_{\mathbf{i}}^{\boldsymbol{\beta}_{\mathbf{i}}(\mathbf{I})} =: \widetilde{B}(h,\mathcal{P}), \qquad h \in (0,1]^d.$$

It remains to choose multi-bandwidth h. To do it it suffices to solve for any $\mathbf{I} \in \mathcal{P}$ the following system of equations:

$$\mathcal{L}_{\mathbf{j}}h_{\mathbf{j}}^{\boldsymbol{\beta}_{\mathbf{j}}(\mathbf{I})} = \mathcal{L}_{\mathbf{i}}h_{\mathbf{i}}^{\boldsymbol{\beta}_{\mathbf{i}}(\mathbf{I})} = \sqrt{\frac{\ln(n)}{nV_{h_{\mathbf{I}}}}}, \qquad \mathbf{i},\mathbf{j} \in \mathbf{I}.$$

The solution is given by

$$h_{\mathbf{i}} = \mathcal{L}_{\mathbf{i}}^{-1/\boldsymbol{\beta}_{\mathbf{i}}(\mathbf{I})}\left(\frac{\mathfrak{L}(\mathbf{I})\ln(n)}{n}\right)^{\gamma_{\mathbf{I}}(\beta,p)/(\boldsymbol{\beta}_{\mathbf{i}}(\mathbf{I})[2+\gamma_{\mathbf{I}}(\beta,p)])}, \qquad \mathfrak{L}(\mathbf{I}) = \prod_{\mathbf{i}\in\mathbf{I}}\mathcal{L}_{\mathbf{i}}^{1/\boldsymbol{\beta}_{\mathbf{i}}(\mathbf{I})}.$$

Here we have also used that $1/\gamma_{\mathbf{I}}(\beta, p) = \sum_{\mathbf{i}\in\mathbf{I}}1/\boldsymbol{\beta}_{\mathbf{i}}(\mathbf{I})$.

It is easy to see that the obtained solution belongs to $\mathcal{H}_n$, and let h^* be its projection on $\overline{\mathcal{H}}_n$, which is, remember, the diadic greed of $\mathcal{H}_n$. It remains to note that $\widetilde{B}(h,\mathcal{P})$ and $\widetilde{B}(h^*,\mathcal{P})$ as well as $V(h,\mathcal{P})$ and $V(h^*,\mathcal{P})$ differ from each other by numerical constants only, and therefore, the assertion of the theorem follows now from Theorem 1. $\square$

APPENDIX

A.1. Proof of Proposition 1. 1^0. Note that $\mathbf{M}(z) = \mathbf{M}(|z|)$; since $\mathbf{M}$ is symmetric this implies

$$\chi_r(y) = n^{-1}\sum_{j=1}^{n}\left[M_r\left(\vec{\rho}(Y_j,y)\right) - \mathbb{E}_g^{(n)}\left\{M_r\left(\vec{\rho}(Y_j,y)\right)\right\}\right],$$

where $\vec{\rho}:\mathbb{R}^s\times\mathbb{R}^s\to\mathbb{R}^s$ is given by $\vec{\rho}(z,z')=(|z_1-z_1'|,\dots,|z_s-z_s'|)$.

We conclude that considered family of random fields obeys the structural assumption introduced in Section 4.4 of Lepski (2012), with $d=s$, $\mathbb{X}_1^d=\bar{\mathbb{X}}_1^d=\mathbb{R}^s$ and $\rho_l:\mathbb{R}\times\mathbb{R}\to\mathbb{R}$, given by $|z-z'|$ for any $l=\overline{1,s}$. It implies in particular that $\mathbb{R}^s$ is equipped with the metric ϱ_s generated by the supremum norm, that is, $\varrho_s=\max_{l=\overline{1,s}}\rho_l$. We remark also that in our case $K(u)=\prod_{l=1}^s\mathbf{M}(u_l)$, $u\in\mathbb{R}^s$, $g\equiv 1$ and $\gamma_l=1$, $l=\overline{1,s}$.

To get the assertion of Proposition 1 we will apply Theorem 9 in Lepski (2012) on $\mathcal{R}_n(s):=[1/n,1]^s$. Note that obviously $\widetilde{\mathcal{R}}_n\subseteq\mathcal{R}_n(s)$. Thus we have to check the assumptions of the latter theorem and to match the notation used in the present paper and in Lepski (2012).

First we note that since $\mathbf{M}$ satisfies Assumption 1, Assumption 9(i) is obviously fulfilled with $L_1=(3s/2)(\mathrm{m}_\infty)^{s-1}L$. Moreover Assumption 9(ii) holds because $g\equiv 1$.

Thus, Assumption 9 is checked.

Consider the collection of closed cubs $\mathbb{B}_{1/2}(\mathbf{j})=\{z\in\mathbb{R}^s:\varrho_s(z,\mathbf{j})\le 1\}$, $\mathbf{j}\in\mathbb{Z}^s$, and let $\mathfrak{E}_{\mathbf{j}}(\delta)$, $\delta>0$ denote the metric entropy of $\mathbb{B}_{1/2}(\mathbf{j})$ measured in the metric ϱ_s.

Obviously $\{\mathbb{B}_{1/2}(\mathbf{j}),\mathbf{j}\in\mathbb{Z}^d\}$ is a countable cover of $\mathbb{R}^s$, and each member of this collection is totally bounded (even compact) subset of $\mathbb{R}^s$. It is easily seen that

$$\operatorname{card}\bigl(\{\mathbf{k}\in\mathbb{Z}^s:\mathbb{B}_{1/2}(\mathbf{j})\cap\mathbb{B}_{1/2}(\mathbf{k})\neq\varnothing\}\bigr)\le 3^s,\qquad\forall\mathbf{j}\in\mathbb{Z}^s.$$

Using the terminology of Lepski (2012) we can say that $\{\mathbb{B}_{1/2}(\mathbf{j}),\mathbf{j}\in\mathbb{Z}^d\}$ is 3^s-totally bounded cover of $\mathbb{R}^s$. Moreover, $\mathfrak{E}_{\mathbf{j}}(\delta)=s[\ln(1/\delta)]_+$ for any $\delta>0$ and any $\mathbf{j}\in\mathbb{Z}^s$. All together, the above allows us to assert that Assumption 7(i) is fulfilled with $\mathbf{I}=\mathbb{Z}^s$, $\mathrm{X}_{\mathbf{j}}=\mathbb{B}_{1/2}(\mathbf{j})$, $N=1.5s$ and $R=1$. It remains to note that Assumption 7(ii) is automatically fulfilled in our case since $g\equiv 1$.

Also we note that for any $\mathbf{j},\mathbf{k}\in\mathbb{Z}^s$ satisfying $\mathbb{B}_{1/2}(\mathbf{j})\cap\mathbb{B}_{1/2}(\mathbf{k})=\varnothing$ one has

$$\inf_{x\in\mathbb{B}_{1/2}(\mathbf{j})}\inf_{y\in\mathbb{B}_{1/2},(\mathbf{k})}\varrho_s(x,y)\ge 1$$

and therefore, Assumption 11 is checked with $\mathfrak{t}=1$. At last we have for any $n\ge 1$,

$$\sup_{r\in\mathcal{R}_n(s)}\sup_{u\notin(0,1]^s}\left|\prod_{l=1}^s\mathbf{M}(u_l/r_l)\right|=0,$$

since $\operatorname{supp}(\mathbf{M})\subseteq[-1/2,1/2]$. Hence, condition (4.24) of Theorem 9 is fulfilled as well, which completes the verification of the assumptions of the theorem.

2^0. Let us match the notation. First, in our case $\mathbf{n}_1=\mathbf{n}_2=n$. Since Y_j, $j\ge 1$, are identically distributed, the quantity denoted $F_{\mathbf{n}_2}(r,\bar{x}^{(d)})$ is given now by $G(r,y)=\int_{\mathbb{R}^s}|M_r(x-y)|\mathbf{g}(x)\,\mathrm{d}x$ and therefore, is independent on n. Here we have taken into account that $\bar{x}^{(d)}\in\mathbb{X}^d=\mathbb{R}^s$.

It is easily seen that

$$G_n:=\sup_{r\in[1/n,1]^s}\|G(r,\cdot)\|_\infty\le\min\bigl[\mathrm{m}_1^s\|\mathbf{g}\|_\infty,\mathrm{m}_\infty^s n^s\bigr]. \tag{A.1}$$

This yields, in particular, that $F_{\mathbf{n}_2} = G_n \le \mathrm{m}_1^s \|\mathbf{g}\|_\infty$ for any $n \ge 1$.

Choosing in Theorem 9 $q = p$, $v = 2p + 2$, $z = 1$ and remembering that $\bar{x}^{(d)} = y$, we have

$$\widehat{\mathcal{U}}^{(v,z,p)}(n, r, \bar{x}^{(d)}) \le \gamma_p(s, \mathrm{m}_\infty)\sqrt{\frac{\bar{G}(r)\ln(n)}{nV_r}}$$

for any $\bar{x}^{(d)} = y \in \mathbb{R}^s$ and any $r \in \widetilde{\mathcal{R}}_n(s) \subseteq \mathcal{R}_n(s)$. To get this assertion we have used that $G_n \le (\mathrm{m}_\infty n)^s$ in view of (A.1).

At last, taking into account that the right-hand side of the latter inequality is independent on y, we deduce from Theorem 9 that for any $p \ge 1$,

$$\mathbb{E}_{\mathbf{g}}^{(n)}\left\{\sup_{r\in\widetilde{\mathcal{R}}_n(s)}\left[\|\chi_r\|_\infty - \gamma_p(s, \mathrm{m}_\infty)\sqrt{\frac{\bar{G}(r)\ln(n)}{nV_r}}\right]\right\}_+^p$$
$$\le c_1(p, s)\left[1 \vee \mathrm{m}_1^s\|\mathbf{g}\|_\infty\right]^{p/2} n^{-p/2} + c_2(p, s)n^{-p},$$

where $c_1(p, s) = 2^{7p/2+5}3^{p+5s+4}\Gamma(p + 1)\pi^p(s, \mathrm{m}_\infty)$ and $c_2(p, s) = 2^{p+1}3^{5s}$. Here we have also used that $G_n \le \mathrm{m}_1^s\|\mathbf{g}\|_\infty$ in view of (A.1), which implies $\widehat{F}_{\mathbf{n}_2} \le 1 \vee \mathrm{m}_1^s\|\mathbf{g}\|_\infty$.

A.2. Proof of Proposition 2. First, noting that $\gamma_p(s, \mathrm{m}_\infty)\sqrt{\mathfrak{a}} = 1/2$ we obtain from Proposition 1 that

$$\mathbb{E}_{\mathbf{g}}^{(n)}\left\{\sup_{r\in\widetilde{\mathcal{R}}_n^{(a)}(s)}\left(\|\chi_r\|_\infty - \frac{1}{2}\sqrt{\bar{G}(r)}\right)\right\}_+^p \le c_n, \tag{A.2}$$

where we have put for brevity $c_n = c_1(p, s)[1 \vee \mathrm{m}_1^s\|\mathbf{g}\|_\infty]^{p/2}n^{-p/2} + c_2(p, s)n^{-p}$. Next, putting $\bar{\chi}_r(y) = \Upsilon_r(y) - \mathbb{E}_{\mathbf{g}}^n\Upsilon_r(y)$ we have in view of (A.2)

$$\mathbb{E}_{\mathbf{g}}^{(n)}\left\{\sup_{r\in\widetilde{\mathcal{R}}_n^{(a)}(s)}\left(\|\bar{\chi}_r\|_\infty - \frac{1}{2}\sqrt{\bar{G}(r)}\right)\right\}_+^p \le c_n. \tag{A.3}$$

To get the latter result we remarked that if $\mathbf{M}$ satisfies Assumption 1, then $|\mathbf{M}|$ satisfies it as well and, therefore, Proposition 1 is applicable to the process $\bar{\chi}_r(\cdot)$. It remains to note that the function $\bar{G}(\cdot)$ is the same for both processes $\chi_r(\cdot)$ and $\bar{\chi}_r(\cdot)$. We also note that

$$G(r) = \sup_{y\in\mathbb{R}^s}\{\mathbb{E}_{\mathbf{g}}^{(n)}\Upsilon_r(y)\}$$

and, therefore, for any $r \in (0, 1]^s$, one has

$$\bar{G}(r) = 1 \vee \left\|\mathbb{E}_{\mathbf{g}}^{(n)}\Upsilon_r\right\|_\infty \le 1 \vee \|\Upsilon_r\|_\infty + \|\bar{\chi}_r\|_\infty, \tag{A.4}$$

where we have used the obvious inequality $|\|x\| \vee \|z\| - \|y\| \vee \|z\|| \le \|x - y\|$ being true for any normed vector space.

Hence, putting $\zeta_n(\mathfrak{a}) = \sup_{r\in\mathcal{R}_n^{(\mathfrak{a})}(s)}[\|\bar{\chi}_r\|_\infty - \frac{1}{2}\sqrt{\bar{G}(r)}]_+$, we obtain for any $r \in \mathcal{R}_n^{(\mathfrak{a})}(s)$

$$\bar{G}(r) \le \tfrac{1}{2}\sqrt{\bar{G}(r)} + 1 \vee \|\Upsilon_r\|_\infty + \zeta_n(\mathfrak{a}).$$

It yields $[\bar{G}(r) - 2(1 \vee \|\Upsilon_r\|_\infty)]_+ \le 2\zeta_n(\mathfrak{a})$, and we have in view of (A.3)

$$\mathbb{E}_{\mathbf{g}}^{(n)}\Big\{\sup_{r\in\mathcal{R}_n^{(\mathfrak{a})}(s)}\big[\bar{G}(r) - 2\big(1 \vee \|\Upsilon_r\|_\infty\big)\big]\Big\}_+^p \le 2^p c_n.$$

Similar to (A.4) we have

$$1 \vee \|\Upsilon_r\|_\infty \le \bar{G}(r) + \|\bar{\chi}_r\|_\infty \le (3/2)\bar{G}(r) + \zeta_n(\mathfrak{a})$$

and therefore, $[1 \vee \|\Upsilon_r\|_\infty - (3/2)\bar{G}(r)]_+ \le \zeta_n(\mathfrak{a})$. Thus we get from (A.3)

$$\mathbb{E}_{\mathbf{g}}^{(n)}\Big\{\sup_{r\in\mathcal{R}_n^{(\mathfrak{a})}(s)}\big[1 \vee \|\Upsilon_r\|_\infty - (3/2)\bar{G}(r)\big]\Big\}_+^p \le c_n.$$

A.3. Proof of Lemma 1. We have in view of Fubini's theorem for any $x_{\mathbf{I}} \in \mathbb{R}^{\mathbf{I}}$

$$\begin{aligned}
s_{h_{\mathbf{I}},\eta_{\mathbf{I}}}^*(x_{\mathbf{I}})
&= \int_{\mathbb{R}^{|\mathbf{I}|}} [K_{h_{\mathbf{I}}} \star K_{\eta_{\mathbf{I}}}](t_{\mathbf{I}} - x_{\mathbf{I}}) f_{\mathbf{I}}(t_{\mathbf{I}})\,\mathrm{d}t_{\mathbf{I}}\\
&= \int_{\mathbb{R}^{|\mathbf{I}|}} \bigg[\int_{\mathbb{R}^{|\mathbf{I}|}} K_{\eta_{\mathbf{I}}}(y_{\mathbf{I}}) K_{h_{\mathbf{I}}}(t_{\mathbf{I}} - x_{\mathbf{I}} - y_{\mathbf{I}})\,\mathrm{d}y_{\mathbf{I}}\bigg] f_{\mathbf{I}}(t_{\mathbf{I}})\,\mathrm{d}t_{\mathbf{I}}\\
&= \int_{\mathbb{R}^{|\mathbf{I}|}} K_{\eta_{\mathbf{I}}}(z_{\mathbf{I}} - x_{\mathbf{I}}) \bigg[\int_{\mathbb{R}^{|\mathbf{I}|}} K_{h_{\mathbf{I}}}(t_{\mathbf{I}} - z_{\mathbf{I}}) f_{\mathbf{I}}(t_{\mathbf{I}})\,\mathrm{d}t_{\mathbf{I}}\bigg]\,\mathrm{d}z_{\mathbf{I}}\\
&= s_{h_{\mathbf{I}}}(x_{\mathbf{I}}) + \int_{\mathbb{R}^{|\mathbf{I}|}} K_{\eta_{\mathbf{I}}}(z_{\mathbf{I}} - x_{\mathbf{I}}) \bigg[\int_{\mathbb{R}^{|\mathbf{I}|}} K_{h_{\mathbf{I}}}(t_{\mathbf{I}} - z_{\mathbf{I}}) \{f_{\mathbf{I}}(t_{\mathbf{I}}) - f_{\mathbf{I}}(z_{\mathbf{I}})\}\,\mathrm{d}t_{\mathbf{I}}\bigg]\,\mathrm{d}z_{\mathbf{I}}.
\end{aligned}$$

Therefore, $\|s_{h_{\mathbf{I}},\eta_{\mathbf{I}}}^* - s_{\eta_{\mathbf{I}}}\|_{\mathbf{I},\infty} \le b_{h_{\mathbf{I}}} \int_{\mathbb{R}^{|\mathbf{I}|}} |K_{\eta_{\mathbf{I}}}(y_{\mathbf{I}})|\,\mathrm{d}y_{\mathbf{I}} \le \mathrm{k}_1^d b_{h_{\mathbf{I}}}$.

A.4. Proof of Lemma 2. The proof of the lemma is completely based on application of Propositions 1–2 and Corollary 1.

PROOF OF (i). Remember that $\zeta(h,\mathcal{P}) = \sup_{\mathbf{I}\in\mathcal{P}}\|\xi_{h_{\mathbf{I}}}\|_{\mathbf{I},\infty}$ and

$$\zeta_n = \sup_{\eta\in\mathcal{H}_n}\sup_{\mathcal{P}\in\mathfrak{P}}\big[\zeta(\eta,\mathcal{P}) - \Lambda A_n(\eta,\mathcal{P})\big]_+.$$

Then we have

$$\begin{aligned}
&\big[\mathbb{E}_f^{(n)}(\zeta_n)^{2q}\big]^{1/(2q)}\\
&\quad = \sum_{\mathcal{P}\in\mathfrak{P}}\sum_{\mathbf{I}\in\mathcal{P}}\bigg(\mathbb{E}_f^{(n)}\bigg\{\sup_{\eta_{\mathbf{I}}\in\mathcal{H}_n^{(\mathfrak{a}_{\mathbf{I}})}(|\mathbf{I}|)}\bigg[\|\xi_{\eta_{\mathbf{I}}}\|_{\mathbf{I},\infty}\\
&\qquad - \gamma_{2q}\big(|\mathbf{I}|,\mathrm{k}_\infty\big)\sqrt{\frac{\bar{s}_n\ln(n)}{nV_{\eta_{\mathbf{I}}}}}\bigg]\bigg\}_+^{2q}\bigg)^{1/(2q)},
\end{aligned} \tag{A.5}$$

where we have put $\mathcal{H}_n^{(\mathfrak{a}_\mathbf{i})}(|\mathbf{I}|) = \{\eta_\mathbf{I} \in (0,1]^{|\mathbf{I}|} : nV_{\eta_\mathbf{I}} \geq \mathfrak{a}_\mathbf{I}^{-1}\ln(n)\}$ and $\mathfrak{a}_\mathbf{I} = [2\gamma_{2q}(\mathbf{I}, \mathrm{k}_\infty)]^{-2}$.

To get the latter result we have used first, that $A_n(\eta, \mathcal{P}) = \sup_{\mathbf{I}\in\mathcal{P}} \sqrt{\frac{\bar{s}_n \ln(n)}{nV_{\eta_\mathbf{I}}}}$ and the trivial inequality $[\sup_i x_i - \sup_i y_i]_+ \leq \sup_i [x_i - y_i]_+$. Next we have used that $\Lambda = \sup_{\mathcal{P}\in\mathfrak{P}} \sup_{\mathbf{I}\in\mathcal{P}} \gamma_{2q}(|\mathbf{I}|, \mathrm{k}_\infty)$. At last we have used that $\eta \in \mathcal{H}_n$ implies $\eta_\mathbf{I} \in \mathcal{H}_n^{(\mathfrak{a}_\mathbf{i})}(|\mathbf{I}|)$ for any $\mathbf{I} \in \mathcal{I}_d$ in view of the definition of $\mathfrak{a}^*$.

Note that for any for any $\mathbf{I} \in \mathcal{I}_d$ and any $\eta_\mathbf{I} \in (0,1]^{|\mathbf{I}|}$

$$\bar{s} \geq 1 \vee \left\| \int_{\mathbb{R}^\mathbf{I}} \left| K_{\eta_\mathbf{I}}(t_\mathbf{I} - \cdot) \right| f_\mathbf{I}(t_\mathbf{I})\, \mathrm{d}t_\mathbf{I} \right\|_{\mathbf{I},\infty} =: \bar{F}_\mathbf{I}(\eta).$$

We conclude that Proposition 1 is applicable with $\chi_r = \xi_{\eta_\mathbf{I}}$, $\mathbf{M} = \mathbf{K}$, $p = 2q$, $s = |\mathbf{I}|$, $\mathfrak{a} = \mathfrak{a}_\mathbf{i}$, $\bar{G} = \bar{F}_\mathbf{I}$, and assertion (i) follows with

$$\mathbf{c}_1(2q, d, \mathbf{K}, \mathbf{f}) = \sum_{\mathcal{P}\in\mathfrak{P}} \sum_{\mathbf{I}\in\mathcal{P}} \left[c_1\bigl(2q, |\mathbf{I}|\bigr)\bigl[1 \vee \mathrm{k}_1^{|\mathbf{I}|}\mathbf{f}\bigr]^q + c_2\bigl(2q, |\mathbf{I}|\bigr)\right]. \qquad \square$$

PROOF OF (ii). Put for any $h \in \mathcal{H}_n$ and $\mathbf{I} \in \mathcal{I}_d$

$$s_\mathbf{I}(h_\mathbf{i}) = \left\| \int_{\mathbb{R}^\mathbf{I}} \left| K_{h_\mathbf{I}}(t_\mathbf{I} - \cdot) \right| f_\mathbf{I}(t_\mathbf{I})\, \mathrm{d}t_\mathbf{I} \right\|_{\mathbf{I},\infty},$$

$$f_{\mathbf{I},n}(h_\mathbf{i}) = \left\| n^{-1} \sum_{i=1}^n \left| K_{h_\mathbf{I}}(X_{\mathbf{I},i} - \cdot) \right| \right\|_{\mathbf{I},\infty}.$$

We have similar to (A.5) $[\bar{s}_n - \bar{\mathbf{f}}_n]_+ \leq \sup_{\mathbf{I}\in\mathcal{I}_d} \sup_{h_\mathbf{I}\in\mathcal{H}_n^{(\mathfrak{a}_\mathbf{i})}(|\mathbf{I}|)} [s_\mathbf{I}(h_\mathbf{I}) - 2\mathbf{f}_{\mathbf{I},n}(h_\mathbf{I})]_+$ and hence

$$\bigl[\mathbb{E}_f^{(n)} [\bar{s}_n - \bar{\mathbf{f}}_n]_+^{2q}\bigr]^{1/(2q)} \leq \sum_{\mathbf{I}\in\mathcal{I}_d} \left(\mathbb{E}_f^{(n)} \Bigl\{ \sup_{h_\mathbf{I}\in\mathcal{H}_n^{(\mathfrak{a}_\mathbf{i})}(|\mathbf{I}|)} \bigl[s_\mathbf{I}(h_\mathbf{I}) - 2\mathbf{f}_{\mathbf{I},n}(h_\mathbf{I})\bigr] \Bigr\}_+^{2q} \right)^{1/(2q)}.$$

Assertion (ii) follows now from the second statement of Proposition 2 with

$$\mathbf{c}_2(2q, d, \mathbf{K}, \mathbf{f}) = \sum_{\mathbf{I}\in\mathcal{I}_d} \left[c_1'\bigl(2q, |\mathbf{I}|\bigr)\bigl[1 \vee \mathrm{k}_1^{|\mathbf{I}|}\mathbf{f}\bigr]^q + c_2'\bigl(2q, |\mathbf{I}|\bigr)\right]. \qquad \square$$

PROOF OF (iii). We have $[\bar{\mathbf{f}}_n - 3\bar{s}_n]_+ \leq 2 \sup_{\mathbf{I}\in\mathcal{I}_d} \sup_{h_\mathbf{I}\in\mathcal{H}_n^{(\mathfrak{a}_\mathbf{i})}(|\mathbf{I}|)} [\mathbf{f}_{\mathbf{I},n}(h_\mathbf{I}) - (3/2)s_\mathbf{I}(h_\mathbf{I})]_+$ and hence

$$\begin{aligned} &\bigl[\mathbb{E}_f^{(n)} [\bar{\mathbf{f}}_n - 3\bar{s}_n]_+^{2q}\bigr]^{1/(2q)} \\ &\quad \leq 2 \sum_{\mathbf{I}\in\mathcal{I}_d} \left(\mathbb{E}_f^{(n)} \Bigl\{ \sup_{h_\mathbf{I}\in\mathcal{H}_n^{(\mathfrak{a}_\mathbf{i})}(|\mathbf{I}|)} \bigl[\mathbf{f}_{\mathbf{I},n}(h_\mathbf{I}) - (3/2)s_\mathbf{I}(h_\mathbf{I})\bigr] \Bigr\}_+^{2q} \right)^{1/(2q)}. \end{aligned}$$

Assertion (iii) follows now from the first assertion of Proposition 2 with

$$\mathbf{c}_3(2q, d, \mathbf{K}, \mathbf{f}) = 2 \sum_{\mathbf{I} \in \mathcal{I}_d} \big[c_1(2q, |\mathbf{I}|)\big[1 \vee \mathrm{k}_1^{|\mathbf{I}|}\mathbf{f}\big]^q + c_2(2q, |\mathbf{I}|)\big]. \qquad \square$$

PROOF OF (iv). Note that

$$\begin{aligned} \bar{\mathfrak{f}}_n &:= 2d^2\mathrm{k}_1^d(\bar{\mathbf{f}}_n)^{d^2/4}\big[\max\{\bar{\mathbf{f}}_n, \mathrm{k}_1^2\mathbf{f}\}\big]^{d-1} \\ &\le \beta\big[(\mathbf{f}_n)^{d^2/4+d-1} + \big(1+\mathrm{k}_1^2\mathbf{f}\big)^{d-1}(\mathbf{f}_n)^{d^2/4} + (\mathbf{f}_n)^{d-1} + \big(1+\mathrm{k}_1^2\mathbf{f}\big)^{d-1}\big], \end{aligned} \tag{A.6}$$

where we have used $\mathrm{k}_1 \ge 1$ and put $\beta = 2d^2\mathrm{k}_1^d 2^{d^2/4+d-1}$. Thus, to get assertion (iv), it suffices to bound from above $\mathbb{E}_f(\mathbf{f}_n)^p$, $p \ge 1$. We obviously have

$$\mathbf{f}_n \le \sum_{\mathbf{I} \in \mathcal{I}_d} \sup_{h_\mathbf{I} \in \mathcal{H}_n^{(\alpha_\mathbf{I})}(|\mathbf{I}|)} \left\| n^{-1} \sum_{i=1}^n \big|K_{h_\mathbf{I}}(X_{\mathbf{I},i} - \cdot)\big| \right\|_{\mathbf{I},\infty},$$

and using Corollary 1 we get for $p \ge 1$,

$$\begin{aligned} &\big[\mathbb{E}_f^{(n)}(\mathbf{f}_n)^p\big]^{1/p} \\ &\quad \le \sum_{\mathbf{I} \in \mathcal{I}_d} \big[1 \vee \mathrm{k}_1^{|\mathbf{I}|}\mathbf{f}\big]^{1/2}\big[\gamma_p(|\mathbf{I}|, \mathrm{k}_\infty) + \big\{c_1\big(p, |\mathbf{I}|\big) + c_2\big(p, |\mathbf{I}|\big)\big\}^{1/p}\big]. \end{aligned} \tag{A.7}$$

Assertion (iv) follows now from (A.6) and (A.7). $\square$

A.5. Proof of Lemma 4. The proof of the lemma is based on the embedding theorem for anisotropic Nikolskii classes which we formulate below.

Let $(\alpha, r) \in (0,\infty)^s \times [1,\infty]^s$ be such that $\varkappa = 1 - \sum_{l=1}^s (\alpha_l r_l)^{-1} > 0$, and let $Q \in (0,\infty)^s$. Then there exists $\mathbf{c} > 0$ completely determined by α, r and s such that

$$\mathbb{N}_{r,s}(\alpha, Q) \subseteq \mathbb{N}_{\infty,s}(\boldsymbol{\alpha}, \mathbf{c}Q), \tag{A.8}$$

where $\boldsymbol{\alpha} = (\boldsymbol{\alpha}_1, \dots, \boldsymbol{\alpha}_s)$, $\boldsymbol{\alpha}_j = \varkappa\alpha_j\varkappa_j^{-1}$ and $\varkappa_j = 1 - \sum_{l=1}^s (r_l^{-1} - r_j^{-1})\alpha_l^{-1}$.

The inclusion (A.8) is a particular case of Theorem 6.9 in Nikol'skiĭ (1977), with $p' = \infty$. We remark that $\mathbb{N}_{\infty,s}(\boldsymbol{\alpha}, Q)$ is anisotropic Hölder class of functions.

Let $\mathbf{E}_i$, $i = \overline{1,s}$ be the family of $s \times s$ matrices where $\mathbf{E}_i = (\mathbf{e}_1, \dots, \mathbf{e}_i, \mathbf{0}, \dots, \mathbf{0})$, and let $\mathbf{E}_0$ be zero matrix. Putting $K(u) = \prod_{l=1}^s \mathbf{K}(u_l)$, $u_l \in \mathbb{R}$, we get for any $\eta \in (0,\infty)^s$ and any $z \in \mathbb{R}^s$,

$$\begin{aligned} \big|\mathcal{B}_{\eta,g}(z)\big| &= \left| \int_{\mathbb{R}^s} K(u)\big[g(z+u\eta) - g(z)\big]\,\mathrm{d}u \right| \\ &\le \sum_{i=1}^s \left| \int_{\mathbb{R}^s} K(u)\big[g(z+\eta\mathbf{E}_i u) - g(z+\eta\mathbf{E}_{i-1}u)\big]\,\mathrm{d}u \right|. \end{aligned}$$

We note that the all components of the vectors $z+\eta\mathbf{E}_i u$ and $z+\eta\mathbf{E}_{i-1}u$ except ith coordinate coincide. Hence using Taylor's expansion we obtain any $\eta\in(0,\infty)^s$ and $z\in\mathbb{R}^s$ in view of (A.8)

$$\left|\int_{\mathbb{R}^s} K(u)\big[g(z+\eta\mathbf{E}_i u)-g(z+\eta\mathbf{E}_{i-1}u)\big]\,\mathrm{d}u\right|$$
$$\le \mathbf{c}Q_i\eta_i^{\boldsymbol{\alpha}_i}\mathrm{k}_1^{s-1}\int_{\mathbb{R}}\big|\mathbf{K}(u)\big||u|^{\boldsymbol{\alpha}_i}\,\mathrm{d}u \le \mathrm{k}_1^s\mathbf{c}Q_i\eta_i^{\boldsymbol{\alpha}_i}.$$

To get the last inequality we have taken into account (3.1) and used that $\mathbf{K}$ is supported on $[-1/2,1/2]$. It is worth mentioning that $\mathbf{c}$ as a function of α is bounded on any bounded domain of $(0,\infty)^s$. Since the right-hand side of the latter inequality is independent of z we come to the assertion of the lemma.

REFERENCES

AKAKPO, N. (2012). Adaptation to anisotropy and inhomogeneity via dyadic piecewise polynomial selection. *Math. Methods Statist.* **21** 1–28. MR2901269

BERTIN, K. (2005). Sharp adaptive estimation in sup-norm for d-dimensional Hölder classes. *Math. Methods Statist.* **14** 267–298. MR2195327

BIRGÉ, L. (2008). Model selection for density estimation with $\mathbb{L}_2$-loss. Available at arXiv:0808.1416v2.

BRETAGNOLLE, J. and HUBER, C. (1979). Estimation des densités: Risque minimax. *Z. Wahrsch. Verw. Gebiete* **47** 119–137. MR0523165

DEVROYE, L. and GYÖRFI, L. (1985). *Nonparametric Density Estimation: The $\mathbb{L}_1$ View*. Wiley, New York. MR0780746

DEVROYE, L. and LUGOSI, G. (1996). A universally acceptable smoothing factor for kernel density estimates. *Ann. Statist.* **24** 2499–2512. MR1425963

DEVROYE, L. and LUGOSI, G. (1997). Nonasymptotic universal smoothing factors, kernel complexity and Yatracos classes. *Ann. Statist.* **25** 2626–2637. MR1604428

DEVROYE, L. and LUGOSI, G. (2001). *Combinatorial Methods in Density Estimation*. Springer, New York. MR1843146

DONOHO, D. L., JOHNSTONE, I. M., KERKYACHARIAN, G. and PICARD, D. (1996). Density estimation by wavelet thresholding. *Ann. Statist.* **24** 508–539. MR1394974

EFROIMOVICH, S. Y. (1986). Non-parametric estimation of the density with unknown smoothness. *Theory Probab. Appl.* **30** 557–568.

EFROMOVICH, S. Y. (2008). Adaptive estimation of and oracle inequalities for probability densities and characteristic functions. *Ann. Statist.* **36** 1127–1155. MR2418652

GACH, D., NICKL, R. and SPOKOINY, V. (2013). Spatially adaptive density estimation by localised Haar projections. *Ann. Inst. Henri Poincaré Probab. Stat.* To appear.

GINÉ, E. and NICKL, R. (2009). An exponential inequality for the distribution function of the kernel density estimator, with applications to adaptive estimation. *Probab. Theory Related Fields* **143** 569–596. MR2475673

GINÉ, E. and NICKL, R. (2010). Adaptive estimation of a distribution function and its density in sup-norm loss by wavelet and spline projections. *Bernoulli* **16** 1137–1163. MR2759172

GOLDENSHLUGER, A. and LEPSKI, O. (2008). Universal pointwise selection rule in multivariate function estimation. *Bernoulli* **14** 1150–1190. MR2543590

GOLDENSHLUGER, A. and LEPSKI, O. (2009). Structural adaptation via $\mathbb{L}_p$-norm oracle inequalities. *Probab. Theory Related Fields* **143** 41–71. MR2449122

GOLDENSHLUGER, A. and LEPSKI, O. (2011). Bandwidth selection in kernel density estimation: Oracle inequalities and adaptive minimax optimality. *Ann. Statist.* **39** 1608–1632. MR2850214

GOLDENSHLUGER, A. and LEPSKI, O. (2012). On adaptive minimax density estimation on $\mathbb{R}^d$. Available at arXiv:1210.1715v1.

GOLUBEV, G. K. (1992). Nonparametric estimation of smooth densities of a distribution in L_2. *Probl. Peredachi Inf.* **28** 52–62. MR1163140

KHASMINSKII, R. and IBRAGIMOV, I. (1990). On density estimation in the view of Kolmogorov's ideas in approximation theory. *Ann. Statist.* **18** 999–1010. MR1062695

IBRAGIMOV, I. A. and KHASMINSKII, R. Z. (1980). An estimate of the density of a distribution. *Zap. Nauchn. Sem. Leningrad. Otdel. Mat. Inst. Steklov.* (*LOMI*) **98** 61–85, 161–162, 166. MR0591862

IBRAGIMOV, I. A. and KHASMINSKII, R. Z. (1981). More on estimation of the density of a distribution. *Zap. Nauchn. Sem. Leningrad. Otdel. Mat. Inst. Steklov.* (*LOMI*) **108** 72–88, 194, 198. MR0629401

JENNRICH, R. I. (1969). Asymptotic properties of non-linear least squares estimators. *Ann. Math. Statist.* **40** 633–643. MR0238419

JUDITSKY, A. and LAMBERT-LACROIX, S. (2004). On minimax density estimation on $\mathbb{R}$. *Bernoulli* **10** 187–220. MR2046772

KERKYACHARIAN, G., LEPSKI, O. and PICARD, D. (2001). Nonlinear estimation in anisotropic multi-index denoising. *Probab. Theory Related Fields* **121** 137–170. MR1863916

KERKYACHARIAN, G., LEPSKI, O. and PICARD, D. (2007). Nonlinear estimation in anisotropic multiindex denoising. Sparse case. *Theory Probab. Appl.* **52** 150–171.

KERKYACHARIAN, G., PICARD, D. and TRIBOULEY, K. (1996). L^p adaptive density estimation. *Bernoulli* **2** 229–247. MR1416864

LEPSKI, O. (2012). Upper functions for positive random functionals. Available at arXiv:1202.6615v1.

LEPSKI, O. V. and LEVIT, B. Y. (1999). Adaptive nonparametric estimation of smooth multivariate functions. *Math. Methods Statist.* **8** 344–370. MR1735470

LEPSKIĬ, O. V. (1991). Asymptotically minimax adaptive estimation. I. Upper bounds. Optimally adaptive estimates. *Theory Probab. Appl.* **36** 682–697.

LEPSKIĬ, O. V. (1992). On problems of adaptive estimation in white Gaussian noise. In *Topics in Nonparametric Estimation. Adv. Soviet Math.* **12** 87–106. Amer. Math. Soc., Providence, RI. MR1191692

MASON, D. M. (2009). Risk bounds for kernel density estimators. *Zap. Nauchn. Sem. S.-Peterburg. Otdel. Mat. Inst. Steklov.* (*POMI*) **363** 66–104, 183. MR2749118

MASSART, P. (2007). *Concentration Inequalities and Model Selection. Lecture Notes in Math.* **1896**. Springer, Berlin. MR2319879

NIKOL'SKIĬ, S. M. (1977). *Priblizhenie Funktsii Mnogikh Peremennykh i Teoremy Vlozheniya*, 2nd ed. "Nauka", Moscow. MR0506247

PARZEN, E. (1962). On estimation of a probability density function and mode. *Ann. Math. Statist.* **33** 1065–1076. MR0143282

REYNAUD-BOURET, P., RIVOIRARD, V. and TULEAU-MALOT, C. (2011). Adaptive density estimation: A curse of support? *J. Statist. Plann. Inference* **141** 115–139. MR2719482

RIGOLLET, P. (2006). Adaptive density estimation using the blockwise Stein method. *Bernoulli* **12** 351–370. MR2218559

RIGOLLET, P. and TSYBAKOV, A. B. (2007). Linear and convex aggregation of density estimators. *Math. Methods Statist.* **16** 260–280. MR2356821

ROSENBLATT, M. (1956). Remarks on some nonparametric estimates of a density function. *Ann. Math. Statist.* **27** 832–837. MR0079873

SAMAROV, A. and TSYBAKOV, A. (2007). Aggregation of density estimators and dimension reduction. In *Advances in Statistical Modeling and Inference*. *Ser. Biostat*. **3** 233–251. World Sci. Publ., Hackensack, NJ. MR2416118

TSYBAKOV, A. B. (1998). Pointwise and sup-norm sharp adaptive estimation of functions on the Sobolev classes. *Ann*. *Statist*. **26** 2420–2469. MR1700239

LABORATOIRE D'ANALYSE, TOPOLOGIE, PROBABILITÉS
UNIVERSITÉ AIX-MARSEILLE
39, RUE F. JOLIOT-CURIE
13453 MARSEILLE
FRANCE
E-MAIL: lepski@cmi.univ-mrs.fr